

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



**Identificação de Fatores Associados à Mudança na
Adesão à Terapêutica Anti-hipertensora**

Sara Margarida Branco Rosa

Trabalho de Projeto
Mestrado em Bioestatística
2015

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



**Identificação de Fatores Associados à Mudança na
Adesão à Terapêutica Anti-hipertensora**

Sara Margarida Branco Rosa

Orientadores: Prof^ª Doutora Maria Helena Mouriño Silva Nunes

Mestre Milene Catarina Coelho Fernandes (IMP)

Trabalho de Projeto
Mestrado em Bioestatística
2015

Resumo

A Hipertensão arterial (HTA) é uma doença crónica que atinge mais de 50% da população portuguesa acima dos 45 anos. O não controlo da HTA está fortemente ligado ao acidente vascular cerebral e ao enfarte agudo do miocárdio, as principais causas de morte em Portugal. A não adesão ao tratamento anti-hipertensivo é um problema significativo e pode causar impacto nos seus resultados.

Apesar de as doenças cardiovasculares serem a maior causa de morte em Portugal, 60-80% dos hipertensos medicados não alcançam a pressão normal. A não adesão e a inércia terapêutica são apontadas como os dois determinantes *major* do não controlo da HTA e estão relacionados com fatores culturais e socioeconómicos.

A prevalência da não adesão é estimada em 25% no geral; 50% em pacientes com hipertensão. A não adesão à medicação anti-hipertensiva está associada a 30% da perda de eficácia do tratamento. O incumprimento do tratamento anti-hipertensivo tem grande impacto negativo na morbilidade, mortalidade e nos custos de saúde.

Este estudo tem como objetivo identificar os fatores que influenciam as mudanças na adesão à medicação. O estudo foi baseado em entrevistas com doentes hipertensos, seguidos entre 2010 e 2011. De forma atingir o objetivo serão utilizados métodos de regressão logística multinomial e de regressão logística binária.

Conclui-se que indivíduos de etnia negra, divorciados, com baixo nível de escolaridade e a viver com várias pessoas na mesma casa têm mais tendência para mudanças negativas, isto é, para se manterem sempre não aderentes ou para mudarem de aderentes para não-aderentes ao longo dos seis meses de estudo.

Palavras-chave: Hipertensão arterial, doenças crónicas, adesão à medicação, mudança na adesão, regressão logística multinomial.

Abstract

Hypertension is a chronic disease that affects more than 50% of the portuguese population above 45 years old. The non-control of hypertension is strongly associated to cerebrovascular accident and acute myocardial infarction, the main causes of death in Portugal. Non-adherence to antihypertensive treatment is a significant problem and may cause an impact on its results.

Although cardiovascular diseases are a major cause of death in Portugal, 60-80% of medicated hypertensive patients do not achieved normal blood pressure. Non-adherence and clinical inertia are considered two major determinants of non-control of hypertension and are related to socioeconomic and cultural elements.

Estimates of the prevalence of non-adherence are the following: 25%, in the general population; 50%, in the hypertensive population. Non-adherence to antihypertensive medication is associated with loss of treatment efficiency in 30% of the cases. The failure of antihypertensive treatment has high negative consequences in morbidity, mortality and health costs.

This study aims at identifying the factors that influence changes in patients' adherence to medication for hypertension. The study relied on interviews with hypertensive patients, wich were carried out between 2010 and 2011. To attain the goal just cited, the multinomial and the binomial logistic regression model has been applied.

We conclude that individuals of black ethnicity, divorced, with a low educational level and with many people at home are more likely to negative changes (to remain always non-adherent to medication or to change from adherent to non-adherent during the six months of study).

Keywords: hypertension, chronical disease, adherence to medication, changes in adherence, multinomial logistic regression.

Índice

Capítulo 1	Introdução	1
1.1	As doenças crônicas e a Hipertensão Arterial	1
1.2	Adesão à terapêutica	4
1.3	Estudo DIMATCH	7
Capítulo 2	Objetivos/Métodos de estudo	11
Capítulo 3	Modelos de regressão	15
3.1	Regressão logística simples	17
3.1.1	Modelo de regressão logística simples	17
3.1.2	Testar a significância dos coeficientes	20
3.1.3	Estimação dos intervalos de confiança	22
3.2	Regressão logística múltipla	24
3.2.1	Modelo de regressão logística múltipla	24
3.2.2	Testar a significância do modelo	27
3.2.3	Estimação dos intervalos de confiança	28
3.3	Como interpretar os modelos estimados?	29
3.3.1	Variável independente em escala nominal com duas categorias	29
3.3.2	Variável independente em escala nominal com mais que duas categorias	30
3.3.3	Variável independente contínua	31
3.4	Estratégias de construção do modelo de regressão logística	32
3.4.1	Seleção <i>stepwise</i> em regressão logística	32
3.5	Ajustamento do modelo de regressão logística	35
3.5.1	Teste de Hosmer e Lemeshow	35
3.5.2	Curva ROC	36
3.6	Regressão logística multinomial	38
3.6.1	Modelo de regressão logística multinomial	38
3.6.2	Interpretação do modelo	41
3.6.3	Construção e Avaliação do modelo	42
Capítulo 4	Descrição das variáveis	43
Capítulo 5	Resultados	55

5.1	Estudo exploratório.....	55
5.2	Regressão logística binária	61
5.2.1	Regressão logística: variável dependente Classificação (Basal)	61
5.2.2	Regressão logística: variável dependente Mudança de adesão (com duas categorias).....	69
5.3	Regressão logística multinomial com variável dependente Mudança de adesão (com três categorias).....	78
5.4	Outros estudos	87
Capítulo 6	Discussão/ Conclusão	89
Capítulo 7	Bibliografia	93
Capítulo 8	Anexos	97
8.1	Códigos utilizados em R.....	97
8.1.1	Leitura e reconhecimento das variáveis no programa R.....	97
8.1.2	Modelos Univariados	98
8.1.3	Modelo logístico com as variáveis escolhidas pelas regressões	101
8.1.4	Modelo logístico utilizando o método Stepwise.....	101
8.1.5	Modelo Multinomial	102

Capítulo 1 Introdução

1.1 As doenças crónicas e a Hipertensão Arterial

Entende-se por doença crónica uma doença que tem uma ou mais das seguintes características: são permanentes, produzem incapacidade/deficiências residuais, são causadas por alterações patológicas irreversíveis, exigem uma formação especial do doente para a reabilitação, ou podem exigir longos períodos de supervisão, observação ou cuidados. (DGS)

As doenças crónicas representam 60% de todas as mortes, a nível global. Entre as causas apontadas para o aparecimento das doenças crónicas encontram-se o envelhecimento da população, a globalização, a urbanização, a alimentação não saudável, a inatividade física, o consumo de tabaco, a hereditariedade e o excesso de peso (WHO). As doenças crónicas ocorrem na maioria em países de baixo e médio rendimento, essencialmente concentradas entre a população mais pobre. Afetam homens e mulheres de forma equitativa e não apenas os mais idosos, sendo que metade dos indivíduos com idades compreendidas entre 30 e 69 anos são portadores deste tipo de doenças (WHO).

O aparecimento de uma doença crónica implica, frequentemente, a modificação dos hábitos de vida e também a necessidade de recorrer a métodos terapêuticos. Por vezes o indivíduo não é capaz de integrar estas mudanças no seu dia-a-dia, resultando daí riscos para a sua saúde (WHO). Por exemplo, alguns estudos demonstraram que muitos indivíduos portadores de doença crónica, como a asma, a hipertensão, a diabetes mellitus e a infeção pelo VIH/SIDA, têm dificuldade em aderir aos esquemas terapêuticos recomendados, o que leva a uma deficiente gestão e fraco controlo da doença. (WHO).

A Hipertensão arterial (HTA) é uma doença crónica determinada por elevados níveis de pressão sanguínea nas artérias, levando o coração a efetuar um esforço adicional para realizar a circulação de sangue através dos vasos sanguíneos. A pressão sanguínea envolve duas medidas, sistólica e diastólica, referentes ao período em que o músculo cardíaco está contraído (sistólica) ou relaxado (diastólica). A pressão normal em repouso situa-se entre os 100 e 140 mmHg para a sistólica e entre 60 e 90 mmHg para a diastólica (WHO).

A importância da HTA está relacionada com (DGS).:

- tem uma prevalência elevada - atinge 46,9% da população portuguesa entre os 35 e 64 anos e 74,9% da população portuguesa acima dos 64 anos;
- é o fator de risco modificável com maior impacto na doença cérebro e cardiovascular, que constituem a maior causa de mortalidade e de morbilidade com grave impacto social em Portugal;
- é difícil de controlar (a maioria dos hipertensos não está controlada);
- tem custos económicos elevados, diretos e indiretos.

O diagnóstico de HTA define-se como a elevação persistente, em várias medições de consultório e em diferentes ocasiões, da pressão arterial sistólica (PAS) igual ou superior a 140 mmHg e/ou da pressão arterial diastólica (PAD) igual ou superior a 90 mmHg. A HTA classifica-se em três graus, correspondendo o grau 1 a hipertensão arterial ligeira (PAS 140-159 mmHg e/ou PAD 90-99 mmHg), o grau 2 a hipertensão arterial moderada (PAS 160-179 mmHg e/ou PAD 100-109 mmHg) e o grau 3 a hipertensão arterial grave (PAS \geq 180 mmHg e/ou PAD \geq 110 mmHg). Esta definição de HTA é válida para pessoas adultas (idade igual ou superior a 18 anos), que não estejam sujeitas a tratamento farmacológico anti-hipertensor e sem outras patologias associadas. Se a PAS e PAD se encontrarem situadas em categorias diferentes de classificação, considera-se a categoria do valor mais elevado (Norma DGS).

A adoção de estilos de vida saudáveis constitui um componente indispensável do tratamento de todas as pessoas com HTA, podendo inclusivamente contribuir para a prevenção da sua ocorrência, em indivíduos suscetíveis. Entre estas medidas encontram-se a redução de peso em indivíduos obesos; adotar uma dieta rica em frutos, vegetais e com baixo teor de gorduras saturadas; redução da ingestão de sal, praticar atividade física; consumir moderadamente bebidas alcoólicas; e cessação do hábito de fumar, importante para a redução do risco cardiovascular (DGS).

Existem grupos particularmente relacionados com a HTA. A idade é um dos principais fatores de risco para a HTA e, de fato, uma grande percentagem de indivíduos com idade superior a 64 anos é hipertensa.(PHYSA) Além disso, alguns estudos revelam que os doentes hipertensos apresentam risco aumentado de deterioração cognitiva e demência e que o controlo da tensão arterial se acompanha de redução significativa desse risco.(DGS) Alguns dados sugerem que a prevalência e gravidade da HTA poderão ser mais elevadas em indivíduos de etnia negra, comparativamente a indivíduos de etnia branca.(DGS) Também a obesidade está fortemente associada a um aumento da prevalência de HTA e de doença cardiovascular. (DGS).

A HTA necessita de terapêutica e vigilância continuada no tempo (Norma DGS). O principal objetivo da terapêutica anti-hipertensora é a prevenção da ocorrência de eventos cardiovasculares e renais e do seu agravamento ou recorrência, pela redução

persistente da tensão arterial para valores considerados normais, com o mínimo de reações adversas e conservando o melhor possível a qualidade de vida (DGS). Por outro lado, a interrupção da terapêutica, absoluta ou intermitente, pode associar-se a um agravamento da situação clínica.

1.2 Adesão à terapêutica

A adesão à terapêutica pode ser definida como um conjunto de comportamentos, tais como tomar a medicação, seguir dietas ou executar mudanças de hábitos de vida que coincidam com o regime terapêutico recomendado por um prestador de cuidados saúde (Almeida et al)(WHO). Assim, a adesão à terapêutica (farmacológica) pode ser definida como o grau de conformidade entre as recomendações do médico prescritor e o comportamento da pessoa relativamente ao regime terapêutico proposto (Haynes et al). A adesão está intimamente ligada à relação dos profissionais de saúde com o doente. Assim sendo, é necessário que se estabeleça uma relação de empatia entre o doente e o médico de forma a criar um clima de confiança muito importante para a aceitação das sugestões dadas relativas ao tratamento (Dias et al).

Existem diferentes fatores que influenciam a não adesão:

- fatores sociais, económicos e culturais como, por exemplo, o nível de escolaridade e a situação profissional;
- fatores relacionados com os serviços e os profissionais de saúde, como o grau de desenvolvimento dos sistemas de saúde, o acesso aos medicamentos e os horários e duração das consultas;
- fatores relacionados com a doença e outras comorbilidades, como a incapacidade física que causa no paciente e o carácter da doença (aguda ou crónica, sintomática ou assintomática);
- fatores relacionados com o tratamento, nos quais se incluem a complexidade, a duração e a realização de alterações na medicação;
- fatores relacionados com o doente, como os conhecimentos, as crenças e perceções do doente relativamente à doença (WHO).

A não adesão pode ser classificada como intencional, quando o doente não cumpre o tratamento apesar de o conhecer, ou não intencional, quando este não compreende a informação prestada pelo profissional de saúde, quando se esquece ou sente dificuldade em gerir a medicação (WHO).

Também se pode mencionar a adesão parcial que se refere a situações em que o doente toma o medicamento na hora errada ou com a dose incorreta. Esta adesão parcial pode também ser intencional e acontece frequentemente na fase mais aguda da doença (quando os doentes se sentem pior têm tendência a aumentar a dose, por outro lado, quando não têm sintomas têm tendência para não tomar). A adesão parcial não intencional é mais comum em idosos e crianças e é causada pelo esquecimento da dose prescrita (causa mais comum), pela confusão nos esquemas de tratamento, rótulo impreciso e/ou incapacidade de abrir as embalagens (Faria, H et al).

Existem vários métodos para avaliar a adesão à terapêutica. O auto-reporte do doente, o método utilizado neste estudo, é o método mais utilizado, normalmente através da aplicação de questionários, com a vantagem de ser barato e de fácil aplicação. No entanto, os doentes poderão apresentar respostas algo enviesadas, sobrestimando a adesão, por existir sempre a necessidade de aceitação social (Dijkstra et al). A opinião do médico é também um método fácil de avaliação. Tendencialmente, os médicos valorizam a adesão total, mas a avaliação efetuada pelo doente também acarreta problemas: doentes que não seguem as recomendações descrevem frequentemente o seu comportamento de forma exata, porém, aqueles que seguem todas as indicações manifestam, por vezes, o seu comportamento de adesão sem exatidão. Outros métodos utilizados são a contagem de comprimidos ou monitorização eletrónica, que consiste num frasco de comprimidos com microchip na respetiva tampa a fim de registar a data e hora em que o frasco é aberto e fechado (WHO).

A adesão à terapêutica é muito importante quando se refere às doenças crónicas (como a HTA), pois estas têm um grande impacto na população. Com o avanço da Medicina, há uma diminuição da taxa de mortalidade e um gradual envelhecimento da população, contribuindo para um aumento da incidência destas doenças. Esta realidade apresenta consequências para a economia mundial, representando cerca de 65% do total de despesas com a saúde em todo o mundo, prevendo-se o seu crescimento até 2020, quer a nível mundial, quer nacional, o que obriga à procura de novas estratégias e alternativas respeitantes aos cuidados de saúde (Machado).

Nos países desenvolvidos, a adesão entre os doentes com doenças crónicas ronda os 50%, sendo que esta percentagem diminui significativamente quando se trata de países em desenvolvimento. A escassez de recursos e as desigualdades sociais no acesso aos cuidados de saúde aumentam substancialmente a não adesão, transformando-a num problema (WHO).

Apesar das doenças cardiovasculares serem a maior causa de morte em Portugal, os indivíduos com hipertensão têm frequentemente a pressão arterial não controlada, sendo estimado que 60-80% dos hipertensos medicados não alcancem a pressão <140/90 mmHg (Macedo et al). A não adesão e a inércia terapêutica são apontadas como os dois determinantes major do não controlo da HTA (Bosworth et al). A não adesão à medicação anti-hipertensiva está associada a 30% da perda de eficácia do tratamento enquanto a adesão está associada a uma diminuição de 38% de risco de problemas cardiovasculares (Denhaerynck et al, 2012)(Márquez-Contreras et al, 2012). Estima-se que 30 a 50% dos doentes, independentemente da doença, do tratamento e do prognóstico, não aderem ao regime terapêutico (Vermeire et al).

A relevância deste estudo está justificada na importância de conseguir encontrar os fatores que influenciam as mudanças na adesão à medicação para a hipertensão arterial ao longo do tempo para que os profissionais de saúde estejam habilitados, com este conhecimento, a adotar algumas alterações quando lidam com diferentes tipos de doentes. O objetivo será diminuir a não adesão à medicação, poupar custos com a saúde, prevenir o aparecimento de novas doenças e baixar a mortalidade associada à hipertensão arterial.

1.3 Estudo DIMATCH

O estudo DIMATCH-HTA (Determinants and Impact of Medication Adherence and Therapeutic Change in the Control of Arterial Hypertension among cohorts of immigrants and non-immigrants at the primary care level) é um estudo observacional, prospetivo em que são seguidos pacientes com hipertensão, medicados para tal e seguidos no centro de saúde. São seguidas duas coortes: uma de imigrantes de primeira geração de países Africanos de língua oficial portuguesa (etnia negra) e uma de não-imigrantes/nativos portugueses (etnia branca). O objetivo será comparar as duas coortes relativamente a adesão à medicação, prevalência da hipertensão controlada, padrões de serviços de saúde, formas de manter a hipertensão controlada, sempre tendo em consideração as diferentes faixas etárias, diferentes comorbilidades, diferentes culturas e diferentes níveis económicos. Pretende-se ainda examinar os potenciais determinantes da adesão à terapêutica, identificar as relações entre adesão, mudança de terapêutica e controlo da pressão arterial, identificar subgrupos em cada coorte com alto risco de pressão arterial não controlada relacionada com não adesão à terapêutica/inércia terapêutica.

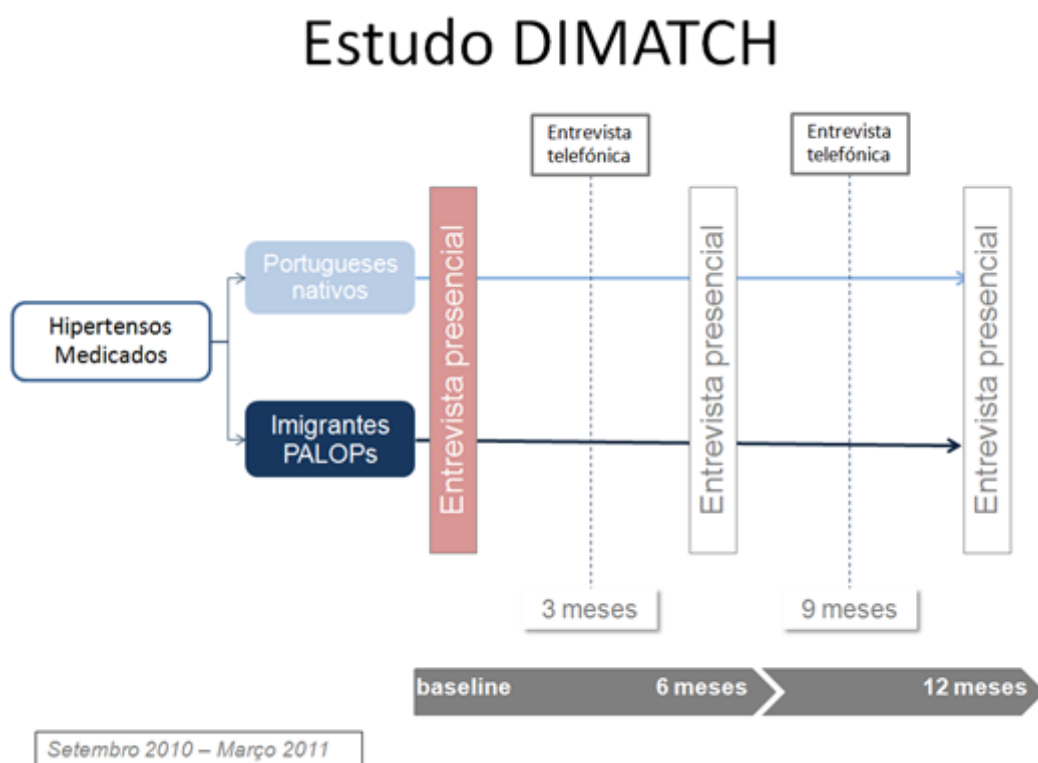
Os critérios utilizados para escolher os pacientes foram os seguintes: ser seguido em centros de saúde da região de Lisboa; ter diagnóstico de Hipertensão arterial (HTA); tomar medicação para a hipertensão arterial; ter consultado o médico há menos de 6 meses; idade entre 40 e 80 anos; ser responsável por tomar a sua medicação (pessoas dependentes de outros foram excluídas); estar contactável, com número de telefone e morada nos ficheiros do centro de saúde e consentir a sua participação no estudo.

O seguimento dos indivíduos foi efetuado durante 12 meses com recolha de informação por entrevista presencial no centro de saúde respetivo (no início do seguimento, designado por momento basal, aos 6 meses e aos 12 meses) e por entrevistas telefónicas (aos 3 meses e 9 meses). As entrevistas iniciais (momento basal) foram realizadas entre setembro de 2010 e março de 2011.

Aplicaram-se vários inquéritos aos pacientes seguidos. Nestes os pacientes indicam os seus dados sociodemográficos (género, idade, nível de escolaridade); se e como tomam os seus medicamentos (sempre, às vezes, ...), quais as estratégias a que recorrem para o controlo da HTA; se consultam o seu médico de família, outro médico, outro tipo de apoio médico, etc. Através destes inquéritos foi possível obter as variáveis que se vão estudar.

Apresenta-se na Figura 1 um esquema da forma como foi conduzido o estudo.

Figura 1 .Esquema do estudo DIMATCH



Os inquéritos começaram por ser aplicados no momento basal a 786 indivíduos, acabando por reduzir este número ao longo da aplicação dos inquéritos. Houve ainda vários indivíduos que não participaram em todas as recolhas de dados (não compareceram ao inquérito presencial aos 6 meses mas compareceram aos 12 meses, por exemplo). Apresento na tabela 1 todas as contagens/percentagens de indivíduos em todos os momentos, assim como uma contagem inicial do número/percentagem de aderentes no momento Basal.

Tabela 1. Número/percentagem de pessoas em cada momento

	Nativos (n(%))	PALOP (n(%))	Total (n)
Basal	449 (57,1%)	337 (42,9%)	786
Basal + 3 meses	371 (58,1%)	268 (41,9%)	639
Basal + 3 + 6 meses	168 (55,3%)	136 (44,7%)	304
Basal + 3 + 6 + 9 meses	151 (58,8%)	106 (41,2%)	257
Basal + 3 + 6 + 9 + 12 meses	99 (63,9%)	56 (36,1%)	155

Número de pessoas em cada coorte no momento basal

Total (n= 786)	Nativos (n= 449)	PALOP (n= 337)
----------------	------------------	----------------

Através das respostas aos inquéritos obtiveram-se 309 variáveis no momento basal, divididas em diversas categorias como variáveis sociodemográficas, relacionadas com os dados sobre o indivíduo (idade, género, estado civil, etc), variáveis clínicas relacionadas com a saúde do indivíduo (IMC, diagnóstico de diabetes, risco cardiovascular, etc), variáveis relacionadas com a frequência dos serviços de saúde (se foi ao médico de família, ao médico privado, ao farmacêutico, etc) e variáveis relacionadas com o controlo da pressão arterial de cada indivíduo (se tem a pressão arterial controlada, medições da pressão arterial sistólica e diastólica, etc).

O estudo DIMATC-HTA foi autorizado pela Comissão de Ética da Faculdade de Medicina de Lisboa, Comissão Nacional de Proteção de Dados e Administração Regional de Saúde de Lisboa e Vale do Tejo. Os Agrupamentos de Centros de Saúde e unidades de cuidados primários aceitaram colaborar no estudo. Todos os participantes assinaram um consentimento informado e esclarecido, apresentado pelas entrevistadoras. O estudo foi desenvolvido pela Unidade de Epidemiologia do Instituto de Medicina preventiva da Faculdade de Medicina de Lisboa (IMP-FML) e co-financiado pela Fundação para Ciência e Tecnologia (FCT), através do projeto PTDC/SAU-ESA/103511/2008, e pela Fundação AstraZeneca.

Capítulo 2 Objetivos/Métodos de estudo

Englobado no estudo DIMATCH-HTA, um estudo prospetivo dos determinantes e do impacto da adesão e da mudança terapêutica no controlo da tensão arterial e da adesão aos anti-hipertensores numa população de hipertensos medicados dos Centros de Saúde de Lisboa, este trabalho procura identificar as variáveis associadas à não adesão à terapêutica, qual a frequência da mudança de adesão ao longo do seguimento (6 meses) e quais as variáveis associadas a essas mudanças. Para cumprir esse objetivo, serão testados vários modelos de regressão (logística binária, logística multinomial), de várias formas para tentar identificar quais as variáveis mais relacionadas com a não adesão à terapêutica.

Se for possível identificar quais as variáveis associadas à não adesão à terapêutica, os médicos e profissionais de saúde poderão tentar alterar a forma como lidam com certos grupos de doentes de forma a diminuir a não adesão.

De forma a cumprir os objetivos propostos será primeiramente efetuado um estudo descritivo separado por classificação, efetuado para Basal-6meses. Esse estudo incluirá os seguintes tópicos:

- Aderente/ Não aderente no Basal, 6 meses
- Aderente / Não aderente ligeiro / Não aderente severo para basal e 6 meses
- Não aderente intencional / Não aderente não intencional para basal e 6 meses
- Mudanças na adesão (Aderente-Aderente, Não Aderente-Aderente, Aderente-Não aderente, Não aderente-Não aderente) para o espaço de tempo entre o Basal e os 6 meses
- Número/percentagem de respostas 'Sim' na MAT (conjunto de perguntas que determinam a classificação como aderente ou não aderente)

Seguidamente será efetuado o mesmo estudo descritivo separado por classificação, efetuado para Basal-12meses. O objetivo será a comparação das proporções quando há uma diferença de 6 meses e 12 meses entre o seguimento dos doentes:

- Aderente/ Não aderente no Basal e 12 meses
- Aderente / Não aderente ligeiro / Não aderente severo para basal e 12 meses
- Não aderente intencional / Não aderente não intencional para basal e 12 meses

- Mudanças na adesão (Aderente-Aderente, Não Aderente-Aderente, Aderente-Não aderente, Não aderente-Não aderente) para o espaço de tempo entre o Basal e os 12 meses
- Número/percentagem de respostas ‘Sim’ na MAT (conjunto de perguntas que determinam a classificação como aderente ou não aderente)

Numa segunda fase, efetuar-se-ão regressões logísticas com variável dependente Classificação utilizando 2 métodos distintos de entrada das variáveis:

- Utilização de diversas regressões univariadas (variável Classificação com todas as variáveis passíveis de entrar no modelo final). Apenas as variáveis cuja regressão obtenha um valor $p < 0,15$ poderão fazer parte do modelo final
- Regressão logística utilizando o método Stepwise existente no programa R.

De seguida tratar-se-á das Mudanças de adesão entre o Basal e 6 meses (não se efetuará para 12 meses porque se teriam poucas observações). Numa primeira fase efetuar-se-á a separação por 2 mudanças, mudança positiva (indivíduos que se mantêm aderentes no Basal e 6 meses e indivíduos que mudam de não aderentes no Basal para Aderentes aos 6 meses) e mudança negativa (indivíduos que se mantêm sempre não aderentes no Basal e 6 meses e indivíduos que mudam de Aderentes no Basal para Não aderentes aos 6 meses). Com esta separação da variável será efetuada uma regressão logística com a variável Mudança na adesão como variável dependente de 2 formas distintas:

- Utilização de diversas regressões bivariadas (variável Mudança com todas as variáveis passíveis de entrar no modelo final). Apenas as variáveis cuja regressão obtenha um valor $p < 0,15$ poderão fazer parte do modelo final
- Regressão logística utilizando o método Stepwise existente no programa R.

De seguida separar-se-á a variável Mudança na adesão em 3 diferentes mudanças: Sempre aderente (Aderente no Basal e 6 meses), Melhor comportamento de adesão (Passa de não aderente no basal para aderente aos 6 meses) e Mau comportamento de adesão (Não aderentes no basal e 6 meses, aderentes no basal e não aderentes aos 6 meses). Para fazer o estudo com a variável mudança de adesão com três categorias como resposta será efetuada uma regressão multinomial utilizando o método empírico:

- Utilização de diversas regressões univariadas (variável Mudança de adesão com todas as variáveis passíveis de entrar no modelo final). Apenas as variáveis cuja regressão obtenha um valor $p < 0,15$ poderão fazer parte do modelo final

Por último será efetuada uma regressão logística univariada com variável dependente classificação e variável independente que indica se é seguido numa USF ou Centro de saúde de forma a tentar perceber se existem diferenças. Será efetuada a mesma regressão com a variável mudança de adesão com 2 categorias.

Capítulo 3 Modelos de regressão

Os modelos de regressão tornaram-se uma componente integral de qualquer análise de dados direcionada para descrever a relação entre variáveis. A variável resposta pode ser uma variável quantitativa (regressão linear) ou pode ser uma variável qualitativa. No caso de a variável resposta ser de natureza categórica (ou qualitativa), os indivíduos que constituem a amostra são classificados de acordo com um certo atributo. Estas variáveis podem ser de natureza nominal, em que não existe nenhuma ordem, ou ordinal, em que existe uma ordem associada à classificação do indivíduo de acordo com o atributo em causa. Se a variável resposta toma apenas dois valores possíveis encontramos-nos no contexto da regressão logística, se toma três ou mais valores será aplicada a regressão multinomial. Neste trabalho serão utilizadas apenas variáveis qualitativas nominais como variável resposta.

O objetivo das análises utilizando modelos de regressão é encontrar o modelo mais simples e apropriado, que faça sentido na aplicação aos dados para descrever a relação entre uma variável resposta (variável dependente) e as várias variáveis explicativas (variáveis independentes ou preditores). As variáveis independentes são também chamadas de covariáveis. Utilizando estes modelos de regressão poderemos prever o valor de uma variável dependente baseando-nos nos valores das variáveis independentes estabelecendo uma relação de natureza estatística.

O modelo de regressão mais comum é o de regressão linear, utilizado quando a variável resposta é contínua. Supondo que existem k covariáveis, o modelo é expresso por

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i, \quad i=1,2,\dots,n$$

sendo n a dimensão da amostra; y_i representa o valor observado; $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ são os parâmetros do modelo de regressão e ε_i é o erro aleatório. Para que se possa aplicar este modelo de regressão existem alguns pressupostos que têm de ser verificados:

- $E[\varepsilon_i]=0$; $\text{Var} [\varepsilon_i]= \sigma^2$; $\text{Cov} [\varepsilon_i, \varepsilon_j]=0, \forall i \neq j$ (condições de Gauss-Markov);
- Os erros são i.i.d. (independentes e identicamente distribuídos) com $\varepsilon_i \sim N(0, \sigma^2)$;

Tendo em consideração que a regressão linear é a mais conhecida e utilizada, tentar-se-á fazer um paralelo entre as regressões utilizadas para este trabalho e a regressão linear, no sentido de entender em que é que diferem.

3.1 Regressão logística simples

3.1.1 Modelo de regressão logística simples

O modelo de regressão linear simples tem como variável resposta uma variável contínua. A equação do modelo de regressão linear simples será

$$Y = E[Y|x] = \beta_0 + \beta_1 x,$$

em que $E[Y|x]$ representa o valor médio da variável aleatória Y condicional a x .

Na regressão logística a variável resposta, a designar por Y , é dicotómica (isto é, só toma dois valores distintos, 0 e 1). Para resolver este problema ter-se-á de modelar a probabilidade de ocorrência de cada um dos acontecimentos, $p = P[Y=1]$ e $1-p = P[Y=0]$, em vez de modelar a variável aleatória Y , utilizando o modelo Binomial.

Considerando um elemento genérico da população com características $(X=x, Y=y)$

$P(Y=y | X=x) \equiv P(Y=y) = p^y (1-p)^{1-y}$ e que

$$Y = \begin{cases} 0, & \text{com probabilidade } 1-p \\ 1, & \text{com probabilidade } p \end{cases}$$

Verifica-se que, a cada indivíduo, temos associado um modelo Binomial, com $Y \sim \text{Binomial}(1, p)$; $E[Y] = p$ e $\text{Var}[Y] = p(1-p)$. Assim sendo, estamos a modelar a resposta média, p , que corresponde à probabilidade de sucesso ($Y=1$, sucesso; $Y=0$, insucesso). Como p é uma probabilidade, então $0 < p < 1$. Assim sendo, há que relacionar a covariável com a variável resposta utilizando uma função que satisfaça a restrição de que o contradomínio seja $]0,1[$. A função logística satisfaz a restrição pretendida e será a utilizada para estabelecer a relação entre X e Y . A função logística tem como expressão

$$f(x) = \frac{e^x}{1+e^x}$$

da qual resulta a forma adaptada que utilizaremos

$$E(Y = y | X = x) \equiv E[Y] = p(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

A transformação de $p(x)$ que utilizaremos para a regressão logística é a transformação logit, obtida através do cálculo da função inversa de $p(x)$. A transformação resultante é definida em termos de $p(x)$ por:

$$g(x) = \ln \left[\frac{p(x)}{1-p(x)} \right] = \beta_0 + \beta_1 x.$$

A função $g(x)$ tem tantas propriedades desejáveis como o modelo de regressão linear: é linear nos seus parâmetros, pode ser continua e pode variar no intervalo $]-\infty, +\infty[$, dependendo do intervalo de variação de x .

Uma segunda diferença entre o modelo de regressão linear e o modelo de regressão logístico está relacionada com a distribuição condicional da variável resposta. Na regressão linear assumimos que as observações da variável resposta condicionais a x , podem ser descritas por $y = E(Y|X=x) + \varepsilon$ em que ε é o erro aleatório e expressa o desvio da observação relativamente à média condicional. Assume-se que ε segue uma distribuição normal, de variância constante. Isto leva a que a distribuição condicional de Y seja normal, com valor médio $E[Y|X=x]$ e variância constante. O mesmo não acontece na regressão logística em que as observações da variável resposta, condicionais a x , serão

$$y = p(x) + \varepsilon.$$

Neste caso ε só pode tomar dois valores: se $y=1$, $\varepsilon=1-p(x)$ com probabilidade $p(x)$; se $y=0$, $\varepsilon=-p(x)$ com probabilidade $1-p(x)$. Assim, ε tem valor médio zero e variância igual a $p(x)[1-p(x)]$. A distribuição condicional da variável aleatória $Y|X=x$ segue uma distribuição Binomial com parâmetro $p(x)$.

Utilizemos uma amostra aleatória de n observações independentes do par (x_i, y_i) , $i=1,2,\dots,n$ e que y_i é codificada para 0 e 1, significando a ausência ou a presença de determinada característica, respetivamente. O ajustamento do modelo logístico exige que sejam estimados β_0 e β_1 , parâmetros desconhecidos do modelo. Os estimadores dos parâmetros, $\hat{\beta}_0$ e $\hat{\beta}_1$ serão obtidos com base na função de verosimilhança, calculando o valor máximo da função. Sendo Y codificado como 0 ou 1, a expressão de $p(x)$ fornece $P(Y=1)$, enquanto que $1-p(x)$ fornece $P(Y=0)$. Assim, para os pares (x_i, y_i) , quando $y_i=1$ a contribuição para a função de verosimilhança é $p(x_i)$; quando $y_i=0$ a contribuição para a função de verosimilhança é $1-p(x_i)$. Assim sendo, obtém-se a seguinte função de verosimilhança:

$$l(\beta) = \prod_{i=1}^n p(x_i)^{y_i} [1 - p(x_i)]^{1-y_i}$$

De forma a ficar matematicamente mais fácil de calcular usamos o logaritmo da função de verosimilhança:

$$L(\beta) = \ln(l(\beta)) = \sum_{i=1}^n \{y_i \ln [p(x_i)] + (1 - y_i) \ln [1 - p(x_i)]\}$$

Para determinar o valor de β que maximiza $L(\beta)$ utilizaremos a derivada da função de verosimilhança em ordem a cada um dos parâmetros, β_0 e β_1

$$\begin{cases} \frac{\partial L}{\partial \beta_0} = \sum_{i=1}^n y_i - \sum_{i=1}^n p(x_i) \\ \frac{\partial L}{\partial \beta_1} = \sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i p(x_i) = \sum_{i=1}^n x_i (y_i - p(x_i)) \end{cases}$$

Assim sendo, os estimadores de máxima verosimilhança obedecem ao sistema de equações :

$$\begin{cases} \sum_{i=1}^n y_i - \sum_{i=1}^n p(x_i) = 0 \\ \sum_{i=1}^n x_i (y_i - p(x_i)) = 0 \end{cases}$$

Os valores de $\beta = (\beta_0, \beta_1)$ obtidos pela resolução do último sistema de equações apresentado são chamados estimadores de máxima verosimilhança e são representados por $\hat{\beta}$ (sempre que estamos perante um estimador, este é representado pelo caracter correspondente associado a ' \wedge '). O estimador de máxima verosimilhança $\hat{\beta}$ tem propriedades importantes, as quais serão descritas no capítulo 3.2.

As equações de máxima verosimilhança são não lineares e necessitam de métodos iterativos para a sua resolução, como o método de Newton-Raphson ou método iterativo dos mínimos quadrados que serão também desenvolvidos no capítulo 3.2.

3.1.2 Testar a significância dos coeficientes

Após a estimação dos coeficientes do modelo ($\hat{\beta}$), o objetivo será avaliar a significância da variável do modelo. Para isso, efetuamos testes estatísticos para determinar se a variável independente está relacionada com a variável dependente.

Na regressão linear, quando testamos a significância de um coeficiente no modelo pretendemos saber se o modelo que contém a variável que estamos a estudar nos diz mais sobre a variável resposta do que o modelo que não contém a variável. Assim, o que podemos fazer é comparar os valores observados da variável resposta com os valores previstos nos dois modelos: com e sem a variável em estudo. Se os valores previstos forem mais exatos quando a variável está presente então esta é significativa.

Para a regressão logística, o objetivo é o mesmo: comparar valores observados da variável resposta com os valores preditos pelos modelos, com e sem a variável independente a estudar. Esta comparação obtém-se através da função log-verosimilhança. Efetua-se a comparação entre as verosimilhanças do modelo ajustado e do modelo saturado utilizando o seguinte teste de razão de verosimilhanças:

$$D_{ajustado} = -2\ln\left[\frac{\text{verosimilhança do modelo ajustado}}{\text{verosimilhança do modelo saturado}}\right] \sim \chi^2_{(n-2)}$$

Consideramos que um valor observado é como um valor predito de um modelo saturado (modelo que contém tantos parâmetros quanto observações, com todas as variáveis no modelo). Sendo a verosimilhança do modelo saturado dada por

$$l(\text{modelo saturado}) = \prod_{i=1}^n y_i^{y_i} \times (1 - y_i)^{1-y_i}$$

e a verosimilhança do modelo ajustado dada por

$$l(\text{modelo ajustado}) = \prod_{i=1}^n \hat{p}(x_i)^{y_i} \times (1 - \hat{p}(x_i))^{1-y_i},$$

com $p(x_i)$ estimado por $\hat{p}(x_i)$, efectua-se a comparação entre as verosimilhanças do modelo ajustado e do modelo saturado utilizando o seguinte teste de razão de verosimilhanças:

$$D_{ajustado} = -2 \ln \frac{\prod_{i=1}^n \hat{p}(x_i)^{y_i} \times (1 - \hat{p}(x_i))^{1-y_i}}{\prod_{i=1}^n y_i^{y_i} \times (1 - y_i)^{1-y_i}} \leftrightarrow$$

$$\leftrightarrow D_{ajustado} = -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{p}(x_i)}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{p}(x_i)}{1 - y_i} \right) \right]$$

O valor D é chamado *deviance* sendo utilizado para calcular a qualidade do ajustamento.

Quando a variável resposta é binária (0 ou 1), a verosimilhança do modelo saturado é sempre igual a 1. Para que tal aconteça é assumido que, se $y_i=0$, o valor obtido por $y_i^{y_i} = 0^0 = 1$ e se $y_i=1$, $(1 - y_i)^{1-y_i} = 0^0 = 1$.

E assim D passa a ser definido apenas como

$$D = -2 \ln(\text{verosimilhança do modelo ajustado})$$

Consideremos agora o modelo nulo, isto é, o modelo só com a constante. Para este modelo temos $E(y_i) = \hat{p}(x_i) = \bar{y}$. Assim sendo, a deviance do modelo nulo é dada por

$$D_{nulo} = -2 \ln \frac{\prod_{i=1}^n \bar{y}^{y_i} \times (1 - \bar{y})^{1-y_i}}{\prod_{i=1}^n y_i^{y_i} \times (1 - y_i)^{1-y_i}} \leftrightarrow$$

$$\leftrightarrow D_{nulo} = -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\bar{y}}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \bar{y}}{1 - y_i} \right) \right]$$

Para testar a significância da variável independente compara-se D calculado com a variável e D com o modelo nulo, utilizando esta fórmula

$$G = D(\text{modelo sem a variável}) - D(\text{modelo com a variável}) \leftrightarrow$$

$$\leftrightarrow G = D_{nulo} - D_{ajustado} \sim \chi_{(1)}^2$$

Como estamos a trabalhar com uma única variável independente, G desempenha a mesma função que a estatística F na regressão linear.

Para calcular a significância dos coeficientes existem ainda outros testes, como o teste de Wald, como utilizado na regressão linear. Na aplicação deste teste compara-se a estimativa de máxima verosimilhança, $\hat{\beta}_1$ com uma estimativa do seu erro padrão. A hipótese de teste é novamente $\beta_1=0$. A estatística de teste é

$$W = \frac{\hat{\beta}_1}{\widehat{SE}(\hat{\beta}_1)} \sim N(0,1)$$

A demonstração dos cálculos utilizados para obter $\widehat{SE}(\hat{\beta}_1)$ pode ser encontrada em Hosmer, Lemeshow e Sturdivant, 2013.

3.1.3 Estimação dos intervalos de confiança

O cálculo dos intervalos de confiança funciona como um bom teste adjunto para a significância do modelo. A construção dos intervalos de confiança é baseada na mesma teoria estatística que se usa para efetuar os testes de significância, mais concretamente no teste de Wald. Assim sendo, o intervalo de confiança 100(1- α)% para β_0 é

$$\hat{\beta}_0 \pm z_{1-\frac{\alpha}{2}} \widehat{SE}(\hat{\beta}_0)$$

e o intervalo de confiança 100(1- α)% para β_1 é

$$\hat{\beta}_1 \pm z_{1-\frac{\alpha}{2}} \widehat{SE}(\hat{\beta}_1)$$

em que $\widehat{SE}(\hat{\beta})$ representa o erro padrão do respetivo estimador e $z_{1-\frac{\alpha}{2}}$ representa o quantil da distribuição N(0,1) correspondente ao nível de significância (α) escolhido.

O intervalo de confiança para o *logit* é dado por

$$\hat{g}(x) \pm z_{1-\frac{\alpha}{2}} \widehat{SE}[\hat{g}(x)]$$

com $\hat{g}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$ e com $\widehat{SE}[\hat{g}(x)]$ obtido através da raiz quadrada da $\text{var}[\hat{g}(x)]$, sendo

$$\text{var}[\hat{g}(x)] = \text{var}(\hat{\beta}_0 + \hat{\beta}_1 x) = \widehat{\text{var}}(\hat{\beta}_0) + x^2 \widehat{\text{var}}(\hat{\beta}_1) + 2x \widehat{\text{cov}}(\hat{\beta}_0, \hat{\beta}_1).$$

A demonstração dos cálculos utilizados para obter $\text{SE}(\hat{\beta}_0)$ e $\widehat{\text{cov}}(\hat{\beta}_0, \hat{\beta}_1)$. pode ser encontrada em *Hosmer, Lemeshow e Sturdivant, 2013*.

3.2 Regressão logística múltipla

3.2.1 Modelo de regressão logística múltipla

Vamos agora generalizar a regressão logística ao caso em que temos mais do que uma variável independente. Assim, consideremos um grupo de p variáveis independentes dadas pelo vetor

$$\mathbf{x}' = (x_1, x_2, \dots, x_p) = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

Para este caso, considera-se que a probabilidade de $Y=1$ condicional aos valores das covariáveis \mathbf{x} é dada por

$$p(\mathbf{x}) = \frac{e^{g(\mathbf{x})}}{1 + e^{g(\mathbf{x})}}.$$

sendo a função logit do modelo de regressão logística múltipla dada por

$$g(\mathbf{x}) = \ln\left(\frac{p(\mathbf{x})}{1-p(\mathbf{x})}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

Se algumas variáveis independentes forem de natureza qualitativa em escala nominal (como por exemplo, a raça, o género, etc), é necessário efetuar uma alteração, transformando-as em variáveis numéricas. Os números utilizados para representar os vários níveis na escala nominal serão apenas identificadores. Esta alteração é feita com recurso a variáveis *dummy*. No geral, se existem k possíveis valores para uma variável, teremos de criar $k-1$ variáveis *dummy*. Suponhamos que a j -ésima variável independente x_j tem k_j níveis. As k_j-1 variáveis *dummy* criadas vão ser denotadas por D_{jl} e os seus coeficientes serão denotados por β_{jl} , com $l=1,2,\dots, k_j-1$, onde

$$D_{jl} = \begin{cases} 1, & \text{se a variável } x_j \text{ tomar o valor associado ao } l - \text{ésimo nível} \\ 0, & \text{caso contrário} \end{cases},$$

$l=1,2,\dots, k_j-1$

Apenas são necessárias k_j-1 variáveis *dummy* e não k_j porque o modelo tem um termo constante.

Assim sendo, o *logit* associado à regressão logística para o caso de a j -ésima variável ser nominal é

$$g(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \dots + \sum_{l=1}^{k_j-1} \beta_{jl} D_{jl} + \beta_p x_p.$$

Assumindo que temos uma amostra com n observações independentes (\mathbf{x}_i, y_i) com $i=1,2,\dots,n$, o objetivo será estimar o vetor $\beta' = (\beta_0, \beta_1, \dots, \beta_p)$ que contém todos os parâmetros desconhecidos do modelo. O método utilizado para estimar os parâmetros é o método da máxima verosimilhança. A função de verosimilhança é idêntica à utilizada para o caso da regressão logística simples ($l(\beta)$), sendo apenas $p(\mathbf{x})$ definido de forma diferente para a regressão logística múltipla. Com base na amostra (\mathbf{x}_i, y_i) , $i=1,2,\dots,n$, podemos escrever a função de verosimilhança como

$$\begin{aligned} \ln l(\beta) &= \ln \left(\prod_{i=1}^n p(\mathbf{x}_i)^{y_i} \times (1 - p(\mathbf{x}_i))^{1-y_i} \right) = \\ &= \sum_{i=1}^n y_i \ln \left(\frac{p(\mathbf{x}_i)}{1-p(\mathbf{x}_i)} \right) + \sum_{i=1}^n \ln (1 - P(\mathbf{x}_i)) = \\ &= \sum_{i=1}^n y_i (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) - \sum_{i=1}^n \ln(1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}) \end{aligned}$$

Assim, ir-se-ão obter $p+1$ equações de verosimilhança através do cálculo da log verosimilhança, respeitantes aos $p+1$ coeficientes. As equações de verosimilhança serão

$$\begin{cases} \frac{\partial L}{\partial \beta_0} = 0 \Leftrightarrow \sum_{i=1}^n y_i - \sum_{i=1}^n p(\mathbf{x}_i) = 0 \\ \frac{\partial L}{\partial \beta_1} = 0 \Leftrightarrow \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n x_{ij} p(\mathbf{x}_i) = 0 \Leftrightarrow \sum_{i=1}^n x_{ij} (y_i - p(\mathbf{x}_i)) = 0, j = 1, 2, \dots, p \end{cases}$$

Uma vez que as equações de verosimilhança não podem ser resolvidas analiticamente, estas serão resolvidas com auxílio de *software*, sendo $\hat{\beta}$ a solução das equações.

Assim, os valores ajustados da regressão logística múltipla são $\hat{p}(x_i)$, calculado utilizando $\hat{\beta}$ e x_i .

Para a regressão logística múltipla é fundamental calcular os erros padrão dos coeficientes. O método para estimar a matriz de variâncias-covariâncias de $\hat{\beta}$ segue também a teoria da máxima verosimilhança. Sabendo que a matriz de variâncias-covariâncias pode ser escrita como

$$\widehat{Var}(\hat{\beta}) = \hat{I}^{-1}(\beta), \text{ onde } \hat{I}(\beta) = - \left. \frac{\partial^2 \ln l(\beta)}{\partial \beta \partial \beta'} \right|_{\beta=\hat{\beta}}$$

Esta teoria indica-nos que os estimadores são obtidos pela matriz de segundas derivadas parciais da função log verosimilhança. Estas derivadas expressam-se da seguinte forma

$$\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = - \sum_{i=1}^n x_{ij}^2 p(x_i)(1 - p(x_i))$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_l} = - \sum_{i=1}^n x_{ij} x_{il} p(x_i)(1 - p(x_i)), j, l=0, 1, 2, \dots, p$$

A matriz $(p+1) \times (p+1)$ que contém os negativos dos termos das equações é chamada $I(\beta)$ e é a matriz de informação de Fisher. As variâncias e covariâncias dos coeficientes estimados são obtidas pelo inverso da matriz, $var(\beta) = I^{-1}(\beta)$. Utiliza-se a expressão $var(\beta_j)$ para denotar o j-ésimo elemento da diagonal da matriz. Os valores dos coeficientes estimados são dados por

$$\widehat{SE}(\beta_j) = [\widehat{var}(\hat{\beta}_j)]^{1/2}, j=1, 2, \dots, p$$

Esta matriz de informação de Fisher é obtida através da matriz \mathbf{X} , designada por matriz das covariáveis, e da matriz \mathbf{V} , a definir à frente. A matriz \mathbf{X} é uma matriz $(n \times (p+1))$, que contem os dados para cada observação, que pode ser escrita como

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}_{n \times (p+1)}$$

e a matriz V é uma matriz $(n \times n)$ diagonal com elemento genérico $\hat{p}(x_i)(1-\hat{p}(x_i))$ que pode ser escrita como

$$\hat{V} = \begin{bmatrix} \hat{p}(x_1)(1-\hat{p}(x_1)) & 0 & \dots & 0 \\ 0 & \hat{p}(x_2)(1-\hat{p}(x_2)) & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \hat{p}(x_n)(1-\hat{p}(x_n)) \end{bmatrix}_{n \times n}$$

Uma forma aproximada de escrever a matriz de informação de Fisher é

$$\hat{I}(\hat{\beta}) = X'VX.$$

3.2.2 Testar a significância do modelo

O primeiro passo para testar a significância do modelo é calcular a significância das variáveis no modelo. O teste da razão de verossimilhanças é feito exatamente como no caso da regressão logística simples (secção 3.1.1), mas, agora para p coeficientes das variáveis independentes. O teste baseia-se na estatística G calculada através da subtração entre a “deviance” do modelo nulo (D_{nulo}) e a “deviance” para o modelo ajustado ($D_{ajustado}$), como já demonstrado para o caso da regressão logística simples (secção 3.1.2). O que muda é o cálculo dos valores ajustados $p(\hat{x}_i)$ que são baseados no vetor $\hat{\beta}$ com $p+1$ parâmetros. Neste teste, considera-se a hipótese nula de que p coeficientes (exceto β_0) são iguais a zero. A estatística de teste G terá distribuição qui-quadrado com p graus de liberdade.

Também é possível aplicar o teste de *Wald*. A estatística de teste para testar que $p+1$ coeficientes são iguais a zero é dada por

$$W = \hat{\beta}'(X'VX)\hat{\beta}$$

com distribuição qui-quadrado com $p+1$ graus de liberdade. Se procurarmos efetuar o teste só para os valores de β associados às variáveis (retirando a constante como no teste da razão de verossimilhanças), basta eliminar o valor $\hat{\beta}_0$ de $\hat{\beta}$ e a sua linha e coluna de $X'VX$.

Para testar a significância para cada parâmetro do modelo em separado aplica-se o teste de Wald para o coeficiente em questão da mesma forma que aplicado para a regressão simples (secção 3.1.2), isto é, utilizando

$$W = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \sim N(0,1)$$

3.2.3 Estimação dos intervalos de confiança

Os intervalos de confiança para cada um dos valores de β são construídos de forma análoga ao caso da regressão logística simples (secção 3.1.3). O intervalo de confiança para o *logit* é mais difícil de calcular. A base é a mesma mas há mais termos envolvidos. O *logit* para o modelo com p covariáveis pode ser expresso por

$$\hat{g}(x) = x' \hat{\beta},$$

com $\hat{\beta}'$ vetor dos $p+1$ coeficientes estimados e x' a representar a constante e o conjunto de valores das p covariáveis do modelo, com $x_0=1$. O intervalo de confiança para o *logit* será

$$\hat{g}(x) \pm z_{1-\frac{\alpha}{2}} \widehat{SE}[\hat{g}(x)]$$

sendo $\widehat{SE}[\hat{g}(x)]$ obtido através da raiz quadrada positiva de $\widehat{var}[\hat{g}(x)]$, a qual é dada por

$$\widehat{var}[\hat{g}(x)] = x'(X'VX)^{-1}x.$$

3.3 Como interpretar os modelos estimados?

Após a estimação do modelo de regressão logística é importante interpretá-lo. Para o fazer, o modelo deve estar devidamente ajustado, de acordo com medidas estatísticas de ajustamento, e as variáveis no modelo devem ser significativas clínica ou estatisticamente. O objetivo quando fazemos a interpretação é tentar entender o que os coeficientes estimados do modelo nos indicam relativamente ao estudo que estamos a efetuar. Os coeficientes das variáveis independentes representam o *slope* (inclinação) da variável dependente por unidade de mudança em cada variável independente. Assim, a interpretação serve para determinar a relação funcional das variáveis independentes e dependente.

Para simplificar, a exposição a realizar nesta secção é baseada no modelo de regressão logística simples.

Um coeficiente associado a uma variável é interpretado como a uma mudança no *logit* correspondente a uma mudança de uma unidade na variável independente que se pretende estudar. Utilizando a notação da regressão logística simples, facilmente generalizada para cada um dos coeficientes da regressão logística múltipla, $g(x) = \beta_0 + \beta_1 x$ e $g(x+1) = \beta_0 + \beta_1 (x+1)$ deste modo, β_1 pode ser escrito como

$$\beta_1 = g(x+1) - g(x)$$

Existem três tipos de variáveis independentes e, de seguida, interpretaremos cada um dos três casos.

3.3.1 Variável independente em escala nominal com duas categorias

Para uma variável independente em escala nominal com duas categorias (0/1), assumindo $x = 1$ obtém-se a função logit $g(1) = \beta_0 + \beta_1$; assumindo $x = 0$, o seu *logit* será $g(0) = \beta_0$. Assim sendo, a diferença entre os logits $g(1)$ e $g(0)$ é dada por:

$$g(1) - g(0) = [\beta_0 + \beta_1] - [\beta_0] = \beta_1.$$

Tendo em conta este resultado, é habitual utilizar uma associação denominada *odds ratio* (OR). Considerando

$$p(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}} ; p(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}},$$

a medida do odds ratio pode ser escrita como

$$OR = \frac{p(1)/[1-p(1)]}{p(0)/[1-p(0)]} = \frac{\frac{\frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}}{1 - \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}}}{\frac{\frac{e^{\beta_0}}{1 + e^{\beta_0}}}{1 - \frac{e^{\beta_0}}{1 + e^{\beta_0}}}}} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_0 + \beta_1 - \beta_0} = e^{\beta_1}$$

Esta medida, *odds ratio* indica-nos o quão mais provável tem de ser a resposta para que esteja mais presente quando $x=1$ e mais ausente quando $x=0$, através da razão entre a probabilidade de sucesso e a probabilidade de insucesso. Por exemplo, se $OR=2$, significa que tem 2 vezes mais probabilidade de estar presente quando $x=1$ que quando $x=0$.

Também podemos calcular o intervalo de confiança $100(1-\alpha)\%$ para o OR que fornece informação adicional do parâmetro

$$e^{\hat{\beta}_1 \pm z_{1-\frac{\alpha}{2}} \widehat{SE}(\hat{\beta}_1)}$$

onde $z_{1-\frac{\alpha}{2}}$ representa o quantil da distribuição $N(0,1)$, para o nível de significância α . Se estivermos a trabalhar com uma variável dicotómica $x = a$ ou $x = b$, pequenas alterações terão de ser efetuadas no cálculo do OR e do respetivo intervalo de confiança, nomeadamente:

$$OR = e^{\beta_1(a-b)}$$

$$IC = e^{[\hat{\beta}_1(a-b) \pm z_{1-\frac{\alpha}{2}} |a-b| \widehat{SE}(\hat{\beta}_1)]}$$

De salientar que no caso de $x = 0$ ou $x = 1$, $a - b = 1$, resultando desta substituição as fórmulas anteriormente utilizadas para o caso $x=0$ ou $x=1$.

3.3.2 Variável independente em escala nominal com mais que duas categorias

Quando se trabalha com uma variável nominal com mais do que duas categorias, utiliza-se uma variável *dummy* para recodificar as diferentes categorias.

Indica-se uma categoria de referência e assim, todas as outras categorias serão comparadas com a de referência, como se estivessemos a trabalhar com uma variável dicotómica. Todos os cálculos de *odds ratio* e intervalos de confiança são semelhantes aos utilizados para as variáveis dicotómicas, apresentados na secção 3.3.1.

3.3.3 Variável independente contínua

Quando há variáveis contínuas, a interpretação do coeficiente depende das unidades de medida da variável. Assumimos que o *logit* é linear para uma covariável contínua. A equação do *logit* é

$$g(x) = \beta_0 + \beta_1 x.$$

Assim

$$g(x+c) - g(x) = c\beta_1$$

determina o logaritmo do *odds ratio* para uma mudança de c unidades em x .

O *odds ratio* associado é calculado por

$$OR(c) = e^{c\beta_1}$$

e o respetivo intervalo de confiança $100(1-\alpha)\%$ é dado por

$$e^{\widehat{c\beta_1} \pm z_{1-\frac{\alpha}{2}} \widehat{cSE(\beta_1)}}.$$

onde $z_{1-\frac{\alpha}{2}}$ representa o quantil da distribuição $N(0,1)$ para o nível de significância α .

3.4 Estratégias de construção do modelo de regressão logística

É natural existirem várias variáveis independentes que devem ser incluídas no modelo. Há que desenvolver uma estratégia e utilizar métodos estatísticos, ou empíricos, para escolher as que se devem usar no modelo final. O objetivo de qualquer método é escolher as variáveis que resultam num melhor modelo.

O critério para incluir uma variável no modelo será diferente de acordo com o problema a tratar. A técnica mais habitual é encontrar o modelo mais parcimonioso que ainda assim explique os dados. A razão para tentarmos minimizar o número de variáveis é obter um modelo mais estável numericamente e mais facilmente generalizável. Apesar das diversas técnicas utilizadas em estatística para construir modelo de regressão, aqui falaremos apenas sobre a técnica *Stepwise*, uma vez que é aquela que se tem revelado mais eficiente do ponto de vista das aplicações.

3.4.1 Seleção *stepwise* em regressão logística

O método de seleção *stepwise* é largamente usado na regressão linear e logística. Muitas vezes existem situações em que não se sabe quais as variáveis a selecionar, por ser um problema cujo estudo é pioneiro e ainda não serem bem entendidas as associações entre as variáveis. Utilizando este tipo de método de seleção das variáveis vamos encontrar as variáveis que estatisticamente se relacionam melhor com a variável resposta.

Qualquer método *stepwise* que inclua ou exclua variáveis é baseado num algoritmo estatístico que avalia a importância das variáveis e as inclui, ou exclui, baseado num critério fixo. Na regressão linear utiliza-se o teste F de *Snedecor* para medir a importância das variáveis; na regressão logística utiliza-se o teste do qui-quadrado para a log-verosimilhança. Em cada passo da aplicação da seleção *stepwise*, a variável mais importante estatisticamente é a que provoca as maiores alterações na log-verosimilhança comparativamente com o modelo que não tem a variável.

Supondo então que temos p possíveis variáveis independentes, todas com possível importância clínica no estudo da variável dependente. Aqui fica a forma de aplicação da seleção *stepwise* por passos.

Passo (0) : Iniciamos o processo com o modelo apenas com a constante e avaliamos a log-verosimilhança (a designar por L_0). De seguida, ajustam-se todos os p possíveis modelos logísticos univariados e comparam-se as respetivas log-verosimilhanças. O

valor da log verosimilhança do modelo que contém a variável x_j no passo zero é designado por $L_j^{(0)}$, com j referente à variável e o expoente (0) referente ao passo. O valor do teste de razão de verosimilhanças entre o modelo com x_j e o modelo só com a constante é dado por $G_j^{(0)} = -2(L_0 - L_j^{(0)})$ e o respetivo valor p é dado por $p_j^{(0)}$.

De salientar que $p_j^{(0)} = P(\chi_{(v)}^2 > G_j^{(0)})$ com $v=1$ se a variável é contínua ou $v=k-1$ se variável é nominal com k categorias.

A variável mais importante é a que apresenta menor valor p , a designar por x_{e1} , a qual é candidata a entrar no modelo. Escolhe-se um valor α_1 fixo para verificar se devemos ou não incluir esta variável no modelo (é recomendado que o valor de α_1 esteja entre 0,15 e 0,20).

Passo (1) : Usamos o modelo já com a variável x_{e1} , além da constante. Seja $L_{e1}^{(1)}$ a log-verosimilhança para este modelo. Para testarmos se mais alguma das restantes $p-1$ variáveis é importante para o modelo, efetuam-se $p-1$ modelos contendo x_{e1} e x_j , com $j=1,2,\dots,p-1$. Para o modelo que contém x_{e1} e x_j seja a log-verosimilhança $L_{e1j}^{(1)}$ e o teste de razão de verosimilhanças $G_j^{(1)} = -2(L_{e1}^{(1)} - L_{e1j}^{(1)})$ e o respetivo valor p , $p_j^{(1)}$. A variável com menor valor p é incluída no modelo, chamar-lhe-emos x_{e2} , e o respetivo valor p , $p_{e2}^{(1)}$. Se este valor $p_{e2}^{(1)}$ for inferior ao valor fixado inicialmente, α_1 , seguimos para o passo 2. Caso contrário, terminamos.

Passo (2) : Ajustamos o modelo com x_{e1} e x_{e2} . A sua log-verosimilhança é dada por $L_{e1e2}^{(2)}$. Vamos efetuar a ‘backward elimination’ para verificar se as variáveis incluídas são verdadeiramente importantes ou podem ser removidas do modelo. Seja $L_{-ej}^{(2)}$ a log-verosimilhança do modelo com x_{ej} retirado. O teste de razão de verosimilhanças deste modelo comparado com o modelo com as duas variáveis é dado por $G_{-ej}^{(2)} = -2(L_{-ej}^{(2)} - L_{e1e2}^{(2)})$ e $p_{-ej}^{(2)}$ o respetivo valor p . Para saber que variável deve ser retirada do modelo é escolhida a variável que, quando removida do modelo obtém um maior valor p . Chamamos a esse valor p $p_{r2}^{(2)}$. Para decidir se a variável x_{r2} deve ser removida, compara-se $p_{r2}^{(2)}$ a um valor α_2 escolhido ($\alpha_2 > \alpha_1$ caso contrário corremos o risco de estar sempre a colocar e retirar a mesma variável do modelo). Se $p_{r2}^{(2)} > \alpha_2$, x_{r2} é removida do modelo, se não mantém-se. Ajusta-se um novo modelo com as variáveis x_{e1} , x_{e2} e as $p-2$ variáveis x_j sobrantes. É verificada a log-verosimilhança de cada modelo comparada com o modelo só com x_{e1} e x_{e2} . Calcula-se razão de verosimilhança e calcula-se o valor p , sendo $p_{e3}^{(2)}$ o modelo com o valor p mais baixo. Se $p_{e3}^{(2)} < \alpha_1$ seguimos para o passo (3), caso contrário terminamos aqui.

Passo (3) : Semelhante ao passo (2). Ajusta-se o modelo incluindo a variável escolhida no passo anterior (passo 1), faz-se a verificação *backward elimination* e de seguida escolhemos outra variável para o modelo. O processo continua até atingirmos o último passo, a designar por Passo (S).

Passo (S) : Acontece quando:

- 1) Todas as variáveis foram incluídas no modelo
- 2) Todas as variáveis no modelo têm valor p de remoção menor que α_2 e as variáveis não incluídas têm valor p maior que α_1 . Neste momento, o modelo contém todas as variáveis importantes relativamente aos critérios α_1 e α_2 .

3.5 Ajustamento do modelo de regressão logística

Consideremos o modelo final, com as covariáveis consideradas significativas para descrever o fenómeno em estudo. Torna-se, então, necessário saber quão efetivamente o modelo descreve a variável resposta. Seja $\mathbf{y}' = (y_1, y_2, \dots, y_n)$ o vetor dos valores observados e seja $\hat{\mathbf{y}}' = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)$ o vetor dos valores estimados. O modelo está bem ajustado se a distância entre \mathbf{y} e $\hat{\mathbf{y}}$ é pequena. Teremos de calcular medidas de distância entre \mathbf{y} e $\hat{\mathbf{y}}$ para saber se está ou não bem ajustado.

No caso da regressão linear é utilizado o coeficiente de determinação, R^2 , para tentar quantificar qual a percentagem de variabilidade explicada da variável resposta pelo modelo. Este coeficiente varia entre 0 e 1, sendo que um R^2 mais próximo de 1 estará melhor ajustado que um modelo com valores mais baixos.

Para a regressão logística a forma usada para calcular R^2 não funciona. Poderão ser calculados vários valores aproximados de R^2 , chamados os pseudo- r^2 . No caso deste trabalho preferi optar por verificar o ajustamento através do teste de Hosmer e Lemeshow e da curva ROC, dos quais se irá falar a seguir.

3.5.1 Teste de Hosmer e Lemeshow

Hosmer e Lemeshow propuseram agrupamentos de acordo com os valores das probabilidades estimadas. Seja J correspondente ao número de valores diferentes observados para x . Suponhamos que $J=n$. Neste caso pensaremos em n colunas correspondentes a n valores de probabilidades estimadas, com a primeira coluna correspondente ao menor valor e a n -ésima coluna correspondente ao maior valor. A estratégia de agrupamento mais usada consiste em efetuar a tabela com diferentes grupos de acordo com os decis da probabilidade estimada. Existirão assim $g=10$ grupos, com o primeiro grupo com $n_1' = n/10$ indivíduos com as mais pequenas probabilidades estimadas e o último grupo com $n_{10}' = n/10$ indivíduos com as maiores probabilidades estimadas. Cada linha estima os valores esperados obtidos somando as probabilidades estimadas de todos os indivíduos no grupo.

Para esta estratégia de agrupamento, a estatística de Hosmer e Lemeshow designada por \hat{C} , é obtida calculando a estatística χ^2 de Pearson para a tabela $g \times 2$ que contém as estatísticas observadas e as frequências estimadas:

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

com n'_k sendo o número total de indivíduos no k -ésimo grupo, c_k é o número de respostas diferentes da covariável no k -ésimo decil,

$$o_k = \sum_{j=1}^{c_k} y_j,$$

é o número de respostas entre os c_k níveis da covariável e

$$\bar{\pi}_k = \sum_{j=1}^{c_k} \frac{m_j \hat{\pi}_j}{n'_k}$$

é a probabilidade estimada média do k -ésimo grupo.

\hat{C} é aproximado a uma distribuição χ^2 com $g-2$ graus de liberdade.

Seguindo o tipo de agrupamento por percentis, obtemos dez grupos, chamados decis de risco. O valor que utilizamos para efetuar o teste é geralmente o valor p e o modelo será bem ajustado se $p > \alpha$.

Para se poder aplicar este teste todas as frequências esperadas terão de ser superiores ou iguais a 5. Se se verificar esta condição conclui-se que existem razões para acreditar que o valor p calculado para este teste é suficientemente preciso para suportar a hipótese de que o modelo está bem ajustado aos dados. Se existir alguma frequência esperada que não cumpra o pressuposto podemos juntar algumas linhas adjacentes de forma a aumentar a frequência, diminuindo assim o número de graus de liberdade.

3.5.2 Curva ROC

A sensibilidade (percentagem de corretamente classificados entre os que têm o atributo ou a taxa de verdadeiros positivos) e a especificidade (percentagem de corretamente classificados nos que não têm o atributo ou a taxa de verdadeiros negativos) servem para classificar um teste como positivo. A relação entre as duas medidas é dada pela área abaixo da curva ROC (*Receiver Operating Characteristic*), a designar por AUC (*Area Under the Curve*). Esta curva originada pela teoria da deteção de sinal mostra como o recetor reage à existência de um ruído na presença de barulho. A

sensibilidade e a especificidade estão relacionadas e variam inversamente. O gráfico da curva ROC mostra como se relacionam estas duas quantidades em cada limiar. Ele calcula a probabilidade de detetar um sinal verdadeiro (sensibilidade) e um falso sinal (1-especificidade). A área abaixo da curva ROC (AUC), que varia entre 0 e 1, dá uma medida da capacidade para discriminar entre os indivíduos que experimentaram o resultado de interesse contra os que não experimentaram.

Consideremos $P(y=1) \geq \alpha$ e $P(y=0) < \alpha$. O valor de α é considerado o ponto de corte, 'cut-of-point' que determina a probabilidade de tomar cada um dos valores da variável resposta (se usar 0.5 significa que mudamos de uma categoria para outra quando $\alpha=0.5$). Quanto maior for o ponto de corte maior é a especificidade do teste e menor a sensibilidade; quanto menor o ponto de corte, maior a sensibilidade e menor a especificidade. Este ponto deverá ser o que maximiza simultaneamente a sensibilidade e a especificidade. Retomando a análise da curva ROC e da AUC, é de salientar que para construir uma curva ROC traça-se um diagrama que represente a sensibilidade em função da proporção de 1- especificidade para um conjunto de pontos de corte. A área abaixo desta curva indica-nos se o modelo distingue com exatidão as duas categorias da variável resposta.

3.6 Regressão logística multinomial

3.6.1 Modelo de regressão logística multinomial

O modelo logístico pode facilmente ser adaptado para uma variável resposta nominal com mais de 2 níveis. O objetivo é modelar as *odds* de escolha de cada nível da variável resposta como função das covariáveis e expressar os resultados através do *odds ratio*, de acordo com a escolha de diferentes níveis. Considerar-se-á apenas a variável resposta de natureza nominal com três níveis (o caso em estudo) mas facilmente se poderia deduzir para mais que três níveis.

Assuma-se que os níveis da variável resposta, a designar por Y, são codificados como 0,1 ou 2. A regressão logística com variável resposta binária é parametrizada através do *logit* $Y=1$ vs $Y=0$. Com 3 níveis, na variável resposta serão necessárias duas funções *logit*. Consideramos como valor de referência $Y=0$ e formam-se funções *logit* $Y=1$ vs $Y=0$ e $Y=2$ vs $Y=0$. Assumindo ainda que existem p covariáveis e um termo constante, denotados por vetor \mathbf{x} de dimensão p+1, onde $x_0 = 1$. As funções logit resultantes são dadas por

$$g_1(\mathbf{x}) = \ln \left[\frac{P(Y=1|\mathbf{x})}{P(Y=0|\mathbf{x})} \right] = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \dots + \beta_{1p}x_p = \mathbf{x}'\boldsymbol{\beta}_1$$

$$g_2(\mathbf{x}) = \ln \left[\frac{P(Y=2|\mathbf{x})}{P(Y=0|\mathbf{x})} \right] = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \dots + \beta_{2p}x_p = \mathbf{x}'\boldsymbol{\beta}_2$$

$$\text{com } \underline{\mathbf{x}}' = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, \quad \boldsymbol{\beta}_1 = \begin{bmatrix} \beta_{10} \\ \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{1p} \end{bmatrix}_{(p+1) \times 1} \quad \text{e} \quad \boldsymbol{\beta}_2 = \begin{bmatrix} \beta_{20} \\ \beta_{21} \\ \beta_{22} \\ \vdots \\ \beta_{2p} \end{bmatrix}_{(p+1) \times 1} \quad ; \beta_{10} = \beta_{20} = 1$$

O vetor dos parâmetros desconhecidos será então escrito como

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{10} & \beta_{20} \\ \beta_{11} & \beta_{21} \\ \vdots & \vdots \\ \beta_{1p} & \beta_{2p} \end{bmatrix}_{(p+1) \times 2} = [\boldsymbol{\beta}_1 \quad \boldsymbol{\beta}_2]$$

Devido à estrutura da função logística, as funções de probabilidade condicional da variável resposta dado o vetor das covariáveis são:

$$P(Y = 1|\mathbf{x}) = \frac{e^{g_1(\mathbf{x})}}{1 + e^{g_1(\mathbf{x})} + e^{g_2(\mathbf{x})}}$$

$$P(Y = 2|\mathbf{x}) = \frac{e^{g_2(\mathbf{x})}}{1 + e^{g_1(\mathbf{x})} + e^{g_2(\mathbf{x})}}$$

$$P(Y = 0|\mathbf{x}) = 1 - P(Y = 1|\mathbf{x}) - P(Y = 2|\mathbf{x}) = \frac{1}{1 + e^{g_1(\mathbf{x})} + e^{g_2(\mathbf{x})}}$$

Uma expressão geral é

$$p_j(\mathbf{x}) \equiv P(Y = j|\mathbf{x}) = \frac{e^{g_j(\mathbf{x})}}{\sum_{i=0}^2 e^{g_i(\mathbf{x})}}, \text{ com } g_0(\mathbf{x}) = 0; \boldsymbol{\beta}_0 = \mathbf{0}.$$

Com base na amostra (\mathbf{x}_i, y_i) , $i=1,2,\dots,n$, construímos a função de verosimilhança

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n P(Y_i = y_i | \mathbf{x}_i)$$

O vetor aleatório $\mathbf{X} = (x_1, x_2, \dots, x_k)'$ segue uma distribuição multinomial, que é uma generalização da distribuição binomial, de parâmetros n e $\mathbf{p} = (p_1, p_2, \dots, p_k)'$, ($p_i > 0$ e $\sum_{i=1}^k p_i = 1$) se

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}, \sum_{i=1}^k x_i = n$$

$$\mathbf{X} \sim \text{Multinomial}(n, p_1, p_2, \dots, p_k)$$

Assim sendo, ter-se-ão de construir variáveis *dummy* para que y_i possa tomar valores 0, 1 ou 2. Consideremos as variáveis

$$Y_{0i} = \begin{cases} 1, & \text{se } y_i = 0 \\ 0, & \text{se } y_i = 1 \text{ ou } y_i = 2 \end{cases}; Y_{1i} = \begin{cases} 1, & \text{se } y_i = 1 \\ 0, & \text{se } y_i = 0 \text{ ou } y_i = 2 \end{cases}$$

$$Y_{2i} = \begin{cases} 1, & \text{se } y_i = 2 \\ 0, & \text{se } y_i = 0 \text{ ou } y_i = 1 \end{cases}, i=1,2,\dots,n$$

Neste contexto,

$$P(Y_i = y_i | \mathbf{x}_i) = P(Y_{0i} = y_{0i}, Y_{1i} = y_{1i}, Y_{2i} = y_{2i} | \mathbf{x}_i)$$

Utilizando esta notação pode escrever-se a função condicional de verosimilhança para uma amostra com n observações independentes como

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n P(Y_i = y_i | \mathbf{x}_i) = \prod_{i=1}^n [p_0(\mathbf{x}_i)^{y_{0i}} p_1(\mathbf{x}_i)^{y_{1i}} p_2(\mathbf{x}_i)^{y_{2i}}]$$

Efetuada o logaritmo e utilizando o facto de $\sum_{j=1}^3 y_{ji} = 1$ vem a log verosimilhança

$$L(\boldsymbol{\beta}) = \ln l(\boldsymbol{\beta}) = \ln \prod_{i=1}^n [p_0(\mathbf{x}_i)^{y_{0i}} p_1(\mathbf{x}_i)^{y_{1i}} p_2(\mathbf{x}_i)^{y_{2i}}] =$$

$$= \sum_{i=1}^n y_{1i} g_1(\mathbf{x}_i) + y_{2i} g_2(\mathbf{x}_i) - \ln(1 + e^{g_1(\mathbf{x}_i)} + e^{g_2(\mathbf{x}_i)})$$

As equações de verosimilhança são determinadas efetuando as primeiras derivadas parciais de $L(\boldsymbol{\beta})$ com respeito a cada um dos $2(p+1)$ parâmetros desconhecidos. A forma geral destas equações é

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_{jk}} = -\sum_{i=1}^n x_{ki} (y_{ji} - p_j(\mathbf{x}_i))$$

para $j = 1, 2; i = 1, 2, \dots, n; e k = 0, 1, 2, \dots, p$ e $x_{0i} = 1 \forall i = 1, \dots, n$.

O estimador de máxima verosimilhança para β , $\hat{\beta}$ é obtido igualando estas equações a zero

$$\frac{\partial L(\beta)}{\partial \beta_{jk}} = 0 \Leftrightarrow \sum_{i=1}^n x_{ki} y_i = \sum_{i=1}^n x_{ki} p_j(x_i)$$

O estimador da matriz de variância-covariância de $\hat{\beta}$ é obtida com base na matriz de informação de Fisher. Torna-se então necessário calcular

$$\frac{\partial^2 L(\beta)}{\partial \beta_{jk} \partial \beta_{jk'}} = - \sum_{i=1}^n x_{k'i} x_{ki} p_j(x_i)(1-p_j(x_i))$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_{jk} \partial \beta_{j'k'}} = \sum_{i=1}^n x_{k'i} x_{ki} p_j(x_i)p_{j'}(x_i)$$

Com $j, j'=1,2$ e $k, k' = 0,1,...,p$.

A matriz de informação de Fisher $I(\hat{\beta})$ é uma matriz $(2p+2) \times (2p+2)$ com elementos que são os simétricos dos valores das equações de segunda derivada avaliados para $\hat{\beta}$. Pode ser escrita como

$$I(\beta)_{(2p+2) \times (2p+2)} = -E \left[\frac{\partial^2 L(\beta)}{\partial \beta \partial \beta'} \right]$$

sendo o respetivo estimador

$$\hat{I}(\hat{\beta}) = -E \left[\frac{\partial^2 L(\beta)}{\partial \beta \partial \beta'} \right] \Bigg|_{\beta=\hat{\beta}}$$

O estimador da matriz de variância-covariância é o inverso da matriz de informação de Fisher, ou seja,

$$\widehat{Var}(\hat{\beta}) = \hat{I}(\hat{\beta})^{-1}$$

3.6.2 Interpretação do modelo

Para simplificar a discussão da estimação e interpretação dos *odds ratio* na regressão logística multinomial vamos generalizar a notação usada no caso binário de

forma a incluir as respostas a serem comparadas e também os valores das covariáveis. Sendo $Y=0$ a referência, o *odds ratio* da categoria $Y=j$ vs $Y=0$ para a covariável $x=a$ vs $x=b$ é

$$OR_j(a, b) = \frac{P(Y = j|x = a)/P(Y = 0|x = a)}{P(Y = j|x = b)/P(Y = 0|x = b)}$$

Neste caso obteremos 2 *odds ratio* diferentes, para a comparação $Y=0$ vs $Y=1$ e para a comparação $Y=0$ vs $Y=2$. A interpretação é igual à efetuada para regressão logística (secção 3.3)

3.6.3 Construção e Avaliação do modelo

As estratégias de construção do modelo são semelhantes às utilizadas para o modelo de regressão logística. O método *Stepwise* está disponível em alguns *softwares* estatísticos e efetua-se da mesma forma (ver secção 3.4).

É um pouco mais complicado de efetuar os cálculos de ajustamento para uma variável resposta com 3 categorias. A estratégia passa por tratar a regressão como duas regressões binárias ($Y=0$ vs $Y=1$ e $Y=0$ vs $Y=2$) e aplicar os métodos de ajustamento já indicados, como a estatística Hosmer e Lemeshow (ver secção 3.5).

Capítulo 4 Descrição das variáveis

Como já foi mencionado anteriormente, o projeto DIMATCH é um projeto muito extenso, com 309 variáveis e 786 observações só no momento Basal. Muitas dessas variáveis têm continuidade nos diferentes momentos (caso mais comum nas entrevistas presenciais aos 6 e 12 meses), outras são criadas apenas nos momentos de contacto telefónico (3 e 9 meses). Com todos estes dados/variáveis foi possível fazer uma grande quantidade de projetos.

Para este projeto, baseado nas mudanças da adesão à terapêutica, o objetivo seria efetuar um estudo longitudinal, ou seja, um estudo ao longo do tempo. Logo como ponto de partida, optou-se apenas por ter em atenção as entrevistas presenciais. Nas entrevistas telefónicas, os entrevistados respondem com menor rigor e as perguntas são ligeiramente diferentes. Em todo o caso, as entrevistas telefónicas foram pensadas mais como uma lembrança para o estudo, para que os entrevistados não passassem 6 meses sem ser recordados do estudo, o que poderia levar a um maior esquecimento da entrevista presencial.

Seria perfeito fazer um seguimento do mesmo indivíduo desde o momento basal, passando pelos 6 meses e terminando nos 12 meses. Primeiro obstáculo: apesar de o momento basal ter as respostas quase todas bem respondidas e sem valores em falta, o mesmo não era esperado para os 6 meses e muito menos para os 12 meses. Esperava-se um menor número de indivíduos a cada entrevista (como verificado no quadro das páginas iniciais), várias respostas em falta/mal respondidas que fariam eliminar o indivíduo da base de dados, várias respostas 'Não sabe/Não responde' que também teriam de ser eliminadas. Segundo obstáculo: Vários indivíduos compareceram apenas aos 6 meses ou apenas aos 12 meses, não aos 3 momentos (apesar de todos terem comparecido no momento Basal). Para tentar perceber qual seria a melhor escolha de seguimento (Basal + 6 meses, Basal + 12 meses ou Basal + 6 meses + 12 meses) seria necessário contar o número de observações válidas para cada um destes 3 casos.

Nesta primeira etapa foi pensado um estudo descritivo utilizando algumas variáveis, uma em particular muito importante: a variável 'Classificação'. A variável Classificação classifica um indivíduo como aderente ou não aderente num determinado momento de acordo com as respostas a um conjunto de questões denominadas MAT: um conjunto de questões adaptadas à população portuguesa a partir da escala de adesão à medicação de Morisky. A escala de adesão de Morisky, com 8 perguntas, foi desenvolvida para obter uma avaliação da adesão reportada pelo próprio doente. Estas escalas são bastante usadas pois poupam-se custos e perdas de

tempo desnecessários. A versão adaptada à população portuguesa tem apenas 7 questões que se encontram na tabela 2:

Tabela 2. Escala de Morisky adaptada para Portugal

	Questão colocada	Hipótese de resposta
1	Já se esqueceu de tomar os seus medicamentos para a hipertensão?	0 – Não 1 – Sim
2	Já tomou os seus medicamentos fora de horas?	0 – Não 1 – Sim
3	Alguma vez parou de tomar os medicamentos para a pressão arterial por se sentir melhor?	0 – Não 1 – Sim
4	Alguma vez parou de tomar os medicamentos para a pressão arterial por se sentir pior?	0 – Não 1 – Sim
5	Alguma vez aumentou a dose da medicação para a pressão arterial por se sentir pior?	0 – Não 1 – Sim
6	Alguma vez deixou de tomar a medicação para a pressão arterial por deixar terminar a medicação?	0 – Não 1 – Sim
7	Alguma vez deixou de tomar a medicação para a pressão arterial por outra indicação que não a do médico?	0 – Não 1 – Sim

Um paciente é considerado '**Não aderente**' se responder 'Sim' a pelo menos uma questão entre estas sete. Podemos ainda separar por 'Não aderente não intencional' se responder 'Sim' às questões 1 ou 2 e 'Não aderente intencional' se responder 'Sim' às questões 3 a 7. Considerou-se ainda o grau de não adesão: 'Não aderente esporádico' se responder 'Sim' apenas a uma questão; 'Não aderente frequente' se responder 'Sim' a mais que uma questão. Quem responde 'Não' a todas as sete questões é classificado como '**Aderente**'.

Para efetuar um estudo de seguimento ao longo do tempo, necessitamos de respostas para a variável 'Classificação' também para os 6 meses e 12 meses, pois um indivíduo poderá mudar de Classificação ao longo do tempo, mudando de 'Aderente' para 'Não-aderente', por exemplo. Por sua vez, para construir a variável 'Classificação', necessitamos de ter uma MAT sem problemas/faltas de resposta/respostas erradas no momento Basal, 6 meses e 12 meses. Tratadas as bases de dados para o momento Basal, 6 meses e 12 meses considerando as variáveis necessárias para o estudo descritivo (MAT), obtivemos três conjuntos de dados cujas dimensões das amostras se encontram na tabela 3.

Tabela 3. Número de pessoas em cada momento do estudo

Basal + 6 meses	Basal + 12 meses	Basal + 6 + 12 meses
424	367	285

O seguimento escolhido foi o Basal + 6 meses (seguimento de meio ano em dois momentos), com 424 observações. O seguimento Basal + 6 + 12 meses (seguimento de um ano em 3 momentos) tem uma perda considerável de indivíduos, que seria ainda maior quanto mais cálculos tivéssemos de efetuar (as regressões efetuadas posteriormente utilizam muitas variáveis o que representa mais problemas/perdas). O seguimento Basal + 12 meses (seguimento de um ano em dois momentos) não apresenta muito interesse pois também tem algumas perdas relativamente ao Basal + 6 meses.

Foi assim efetuado um estudo descritivo a propósito da variável 'Classificação' nos vários momentos e na variável '**Mudança**' de adesão construída à custa da variável 'Classificação'. A forma de construção desta variável 'Mudança' foi a seguinte:

$$\text{Mudança} = (2 \times \text{ClassificaçãoBasal}) + \text{Classificação6meses}$$

Se um indivíduo é classificado como 'Aderente' nos dois momentos, toma o valor '**Sempre aderente**' (0); se um indivíduo muda de 'Aderente' no primeiro momento para 'Não Aderente' no segundo momento, é classificado como '**Mudança para pior**' (1); se um indivíduo muda de 'Não aderente' no primeiro momento para 'Aderente' no segundo momento, classifica-se como '**Mudança para melhor**' (2); se um indivíduo é 'Não aderente' nos dois momentos, classifica-se como '**Sempre não aderente**' (3).

Em resumo, as variáveis usadas para efetuar o estudo descritivo relativamente à Adesão em cada momento e à mudança encontram-se na tabela 4.

Tabela 4. Variáveis para o estudo descritivo

Variável	Descrição	Hipótese de resposta
P1	Questão nº 1 da MAT	0 – Não 1 – Sim
P2	Questão nº 2 da MAT	0 – Não 1 – Sim
P3	Questão nº 3 da MAT	0 – Não 1 – Sim
P4	Questão nº 4 da MAT	0 – Não 1 – Sim
P5	Questão nº 5 da MAT	0 – Não 1 – Sim
P6	Questão nº 6 da MAT	0 – Não 1 – Sim
P7	Questão nº 7 da MAT	0 – Não 1 – Sim
Soma	Soma do número de ‘Sim’ da MAT	
Classificacao	Classificação de um individuo relativamente à adesão à medicação anti-hipertensiva	0 – Aderente 1 – Não aderente
Grau	Avaliação do Grau de não adesão à medicação anti-hipertensiva	0 – Aderente 1 – Não aderente esporádico 2 – Não aderente frequente
Intencao	Avaliação da Intencionalidade ou não da não adesão	0 – Não intencional 1 – Intencional
Mud_adesao	Avaliação da mudança ou não na adesão ao longo de dois momentos	0 – Sempre Aderente 1 – Mudança para pior 2 – Mudança para melhor 3 – Sempre não aderente

Todas as variáveis excepto a Mudança na adesão (combinação de dois momentos) têm a sua versão em dois momentos distintos (Basal e 6 meses). Apesar de ter sido escolhido o seguimento Basal + 6 meses foi também efetuado o estudo para Basal + 12 meses, poderá ser útil para verificar as diferenças num seguimento de 6 meses e seguimento de um ano. Todos os cálculos descritivos foram efetuados com auxílio do software IBM SPSS Statistics 20.

Este projeto foi sempre pensado para, de acordo com o original, ser separado por Etnias (Etnia branca vs Etnia negra). No entanto, no decorrer do trabalho, concluiu-se que seria muito extenso e confuso apresentar todas as opções (Etnia Negra, Etnia Branca e Todos sem separação), optando apenas por apresentar os cálculos sem separação por Etnias, sendo a Etnia considerada uma simples variável.

Após a escolha do seguimento Basal + 6 meses, o próximo passo seria efetuar uma regressão logística com variável resposta 'Classificação' e outra com variável resposta Mudança. Começou por ser pensado um modelo que incluiria variáveis do momento Basal e variáveis semelhantes aos 6 meses. Esta tarefa mostrou-se muito complicada pois muitas perguntas foram colocadas de forma diferente, havia muitas perguntas sem resposta, muitos erros. Por outro lado algumas das variáveis mais importantes são iguais nos dois momentos (Sexo, Idade, Etnia, etc). Foi então decidido usar como covariáveis apenas as variáveis obtidas no Basal para tentar explicar as mudanças que ocorrem na adesão à medicação anti-hipertensiva. Como foram utilizadas mais variáveis houve também mais perdas de observações, tal como era esperado. Muitas variáveis tinham respostas erradas (por exemplo: idade = 999), tinham respostas em branco ou respostas de difícil interpretação (as respostas 'Outras' e as respostas 'NS/NR', impossíveis de colocar em alguma categoria). No final, com a base de dados limpa e sem problemas, obtivemos apenas **285** observações, confirmando que foi uma boa opção deixar para trás o seguimento Basal + 6 meses + 12 meses que ficaria com um número de observações na ordem das 100 pessoas.

As variáveis usadas para a regressão logística foram variáveis já testadas e pensadas para este tipo de problemas, usadas também na literatura e escolhidas por essa razão. Apresentar-se-ão primeiro por grupo de variáveis e seguidamente com uma tabela explicativa do tipo e resposta da mesma. (tabela 5)

Variáveis sociodemográficas: Idade, Género, EstadoCivil, nAnosEscola, nPessoasCasa, cOcupacao, cFinancDific.

Variáveis Clínicas: IMC_3categorias, DiagnDiabetes, RCV, cColesterol, cCtrlHTA, cAjudaHTA, FreqCod, idadeHTA, Num_Meds_HTA, Num_Meds_basal.

Variáveis de serviços de saúde: cRecMedFamiliaUltAno, cRecEspecialistaUltAno, cRecMedPrivadoUltAno, cRecUrgenciaUltAno, cRecAtendCSUltAno, cRecFarmaceuticoUltAno, cSegFarmaciaUltAno, cEnfermUlt.

Variáveis relacionadas com o controlo da pressão arterial: PA_Control_2cat, MediaPAS, MediaPAD.

Tabela 5. Variáveis utilizadas na regressão logística

Variável	Classificação	Resposta
Idade (Idade)	Quantitativa Contínua	
Género (Genero)	Qualitativa Nominal	0-Feminino 1-Masculino
Etnia (Etnia)	Qualitativa Nominal	0-Etnia Branca 1-Etnia negra
Ocupação (cOcupacao)	Qualitativa Nominal	1-Trabalhador/Estudante/Dona de casa 2-Desempregado 3-Reformado/Doente/ Permanentemente incapacitado
Estado Civil (EstadoCivil)	Qualitativa Nominal	1-Solteiro 2-Casado ou em união de facto 3-Divorciado ou separado 4-Viúvo
Anos de escolaridade completos (nAnosEscola)	Quantitativa Discreta	
Vive com quantas pessoas (nPessoasCasa)	Quantitativa Discreta	
Dificuldade financeira para comprar medicamentos (cFinancDific)	Qualitativa Ordinal	1-Não foi nada difícil (sem dificuldade) 2-Foi um pouco difícil (Alguma dificuldade) 3-Muito difícil (Muita dificuldade)
Idade de diagnóstico de HTA (idadeHTA)	Quantitativa Contínua	
Número de medicamentos (Num_Meds_basal)	Quantitativa Contínua	
Número de Medicamentos para HTA (Num_Meds_HTA)	Quantitativa Contínua	
IMC (IMC_3categorias)	Qualitativa Ordinal	1-Baixo peso/normal 2-Excesso de peso 3-Obesidade
Tem diabetes (DiagnDiabetes)	Qualitativa Nominal	0-Não 1-Sim
Tem colesterol (cColesterol)	Qualitativa Nominal	1-Sim 2-Não
Risco cardiovascular (RCV)	Qualitativa Nominal	1-Sem RCV 2-Com RCV
Considera sua HTA controlada (cCtrlHTA)	Qualitativa Nominal	1-Sim 2-Não
Tem ajuda para controlar HTA (cAjudaHTA)	Qualitativa Nominal	1-Sim 2-Não

Tabela 5. Variáveis utilizadas na regressão logística (cont.)

Variável	Classificação	Hipótese de resposta
Frequência da medição da HTA (FreqCod)	Qualitativa Nominal	1-Medição regular 2-Medição pouco regular 3-Outras
Recorreu ao médico de família no último ano (cRecMedFamiliaUltAno)	Qualitativa Nominal	1-Sim 2-Não
Recorreu ao especialista no hospital no último ano (cRecEspecialistaUltAno)	Qualitativa Nominal	1-Sim 2-Não
Recorreu ao médico privado no último ano (cRecMedPrivadoUltAno)	Qualitativa Nominal	1-Sim 2-Não
Recorreu ao serviço de urgência no último ano (cRecUrgenciaUltAno)	Qualitativa Nominal	1-Sim 2-Não
Recorreu ao atendimento complementar no CS no último ano (cRecAtendCSUltAno)	Qualitativa Nominal	1-Sim 2-Não
Recorreu ao farmacêutico no último ano (cRecFarmacuticoUltAno)	Qualitativa Nominal	1-Sim 2-Não
É seguido de forma regular na farmácia no último ano (cSegFarmaciaUltAno)	Qualitativa Nominal	1-Sim 2-Não
Recorreu ao enfermeiro no último ano (cEnfermUlt)	Qualitativa Nominal	1-Sim 2-Não
Controlo TA 140/90 (PA_Control_2cat)	Qualitativa Nominal	1- <i>Controlled</i> 2- <i>Uncontrolled</i>
Média PASistólica (MediaPAS)	Quantitativa Contínua	
Média PADiastólica (MediaPAD)	Quantitativa Contínua	
Classificacao (Classificacao)	Qualitativa Nominal	0-Aderente 1-Não Aderente
Mudanças adesão (Mud_adesao)	Qualitativa Nominal	0-Sempre aderente 1-Mudança para pior 2-Mudança para melhor 3-Sempre não aderente

Muitas variáveis pensadas inicialmente para fazer parte do estudo foram sendo recodificadas/ retiradas por variados problemas:

Ocupação A variável Ocupação foi recodificada. Tinha inicialmente 10 categorias que foram recodificadas para 3, de acordo com a tabela 6

Tabela 6. Recodificação da variável Ocupação

Categorias existentes	Recodificação
1-Exerce uma profissão 2-Trabalhador/Estudante 3-Tarefas domésticas	1-Trabalhador/ Estudante/Dona de casa
4-Desempregado <1ano 5-Desempregado 1+ anos	2-Desempregado
6-Reformado 7-Permanentemente Incapacitado 8-Doente com duração ≥ 3 meses 9-Doente com duração < 3 meses	3-Reformado/Doente/ Permanentemente incapacitado
10- Outra Situação	Eliminada

Estas recodificações são efetuadas porque existem categorias com um número de observações muito reduzido. Assim torna-se possível comparar as diferentes categorias e haverá menos problemas com as variáveis.

Tempo de residência em Portugal: Esta variável foi pensada para trabalhar por etnias. Indica o tempo de residência em Portugal, faz mais sentido para os imigrantes Africanos. Não tinha praticamente nenhuma resposta para a Etnia Branca. No princípio tentou-se colocar a idade como variável proxy para o tempo de residência da etnia branca (tendo em consideração que nasceram em Portugal, vivem cá desde que nasceram). Rapidamente apareceram os problemas de multicolinearidade e a variável foi retirada.

Fala Português em casa: Retirada pela mesma razão da variável anterior. Pensada para a separação por etnia. A Etnia Branca só apresentava respostas “sim”, pelo que não fazia sentido quando utilizada sem separação por etnias.

Índice de Graffar: Esta variável indica valores para o índice de Graffar. O índice de Graffar é um índice utilizado para determinar a classificação social e económica de determinada pessoa. É baseada em cinco fatores: Profissão, anos de escolaridade,

fontes de rendimento familiar, conforto do alojamento e aspeto do bairro onde habita. A pontuação final corresponde à classe social:

- 1- Nível Social e Económico Alto
- 2- Nível Social e Económico Médio/Alto
- 3- Nível Social e Económico Médio
- 4- Nível Social e Económico Médio/Baixo
- 5- Nível Social e Económico Baixo

É uma variável com algum interesse mas apresentava uma grande correlação com a variável 'Número de anos na escola', a qual contribuiu para a sua construção. Depois de uma revisão da literatura foi decidido manter a variável 'Número de anos na escola', em detrimento do índice de Graffar.

IMC: A variável IMC tinha inicialmente 4 categorias que foram recodificadas de acordo com a tabela 7.

Tabela 7. Recodificação da variável IMC

Categorias existentes	Recodificação
1-Baixo peso 2-Peso Normal	1-Baixo Peso/Normal
3-Excesso de peso	2-Excesso de peso
4-Obesidade	3- Obesidade

FreqCod: A variável FreqCod foi uma recodificação da variável cFreqHTA que indica a frequência da medição da Hipertensão arterial. Tinha inicialmente 7 categorias, algumas com pouquíssimas observações e categorias sem nenhuma observação, o que trazia alguns problemas. A codificação inicial e a recodificação encontram-se na tabela 8.

Tabela 8. Recodificação da variável FreqCod/cFreqHTA

Categorias existentes	Recodificação
1-Uma ou mais vezes por semana 2- Uma ou mais vezes por mês	1-Medicação Regular
3- Três em três meses 4- Seis em seis meses 5-Uma vez por ano	2-Medicação pouco regular
6-Nunca 7-Outra	3-Outras

Variáveis de serviços de saúde: No início do estudo foi pensado incluir todas as variáveis relativas à ida ao médico/serviço de saúde. Estavam disponíveis dois grupos de variáveis: as variáveis que inquiriam se um indivíduo foi ao médico/farmácia/enfermeiro no último ano e as variáveis que inquiriam o mesmo sem questionar quando o fez. Logo nos deparamos com um problema pois quando alguém responde que recorreu ao serviço no último ano também responderá o mesmo para as variáveis que não questionam quando recorreu. Não só eram muitas variáveis que perguntavam o mesmo como ainda se corria o risco de multicolinearidade entre as variáveis. Assim, foi decidido manter apenas as variáveis relativas ao último ano.

MAP: Incluída inicialmente, indicava a pressão arterial média utilizando a pressão arterial sistólica e diastólica. Foi retirada por risco de multicolinearidade.

Algumas variáveis foram excluídas logo durante o tratamento da base de dados porque tinham muitos valores omissos que levariam a muitas observações retiradas. Como exemplo indicamos a variável que questionava o dinheiro gasto com medicamentos (229 não respostas, em 424).

Para efetuar uma regressão logística (2 categorias) e uma regressão multinomial (3 categorias), a variável mudança terá de ser também ela recodificada das seguintes formas que apresentam os casos mais significativos da Mudança de adesão (tabela 9).

Tabela 9. Recodificação da variável Mudança para duas e três categorias**Mudança (duas categorias)**

Categorias existentes	Recodificação
0- Sempre aderente 2- Mudança para melhor	0 – Mudança positiva
1-Mudança para pior 3-Sempre não aderente	1 – Mudança Negativa

Mudança (três categorias)

Categorias existentes	Recodificação
0- Sempre aderente	0 – Sempre aderente
2- Mudança para melhor	1 – Melhor comportamento de adesão
1-Mudança para pior 3-Sempre não aderente	2- Mudança/manutenção de não adesão

Classificacao Pensou-se utilizar a variável Classificação como co-variável para a regressão logística utilizando a variável dependente Mudança. Acabou por não ser usada por ser parte integrante da variável Mudança existindo assim hipótese de problemas de multicolinearidade.

Num procedimento seguinte foi ainda necessário criar a variável CS_USF_Cod para tentar entender se existem diferenças nas Mudanças de adesão de acordo com a unidade de saúde consultada; se Centro de Saúde (CS, $y=0$) ou Unidade de saúde familiar (USF, $y=1$). Isto porque as USF prestam cuidados de saúde primários personalizados, num quadro de contratualização interna, envolvendo objetivos de acessibilidade, adequação, efetividade, eficiência e qualidade, o que poderia levar a um melhor desempenho dos prestadores de saúde. É uma variável simples, nominal,

que atribui o valor 0 se o paciente é consultado no Centro de saúde e o valor 1 se o paciente é consultado na Unidade de saúde familiar.

Capítulo 5 Resultados

5.1 Estudo exploratório

A primeira abordagem no tratamento dos dados foi efetuar um pequeno estudo exploratório relativo à Classificação de cada indivíduo como 'Aderente' ou 'Não Aderente' de forma a tentar compreender a percentagem/número de pacientes que mudam a sua Classificação e o seu comportamento. Relembra-se que estes testes foram aplicados numa base de dados com **424 indivíduos**. Começamos com uma comparação no seguimento mais importante deste estudo: Basal-6meses. Todos os cálculos foram obtidos com auxílio do software IBM SPSS Statistics 20.

Tabela 10. Classificação no momento Basal e 6 meses

	Basal (n(%))	6 Meses (n(%))
Aderente	175 (41,3%)	237 (55,9%)
Não Aderente	249 (58,7%)	187 (44,1%)
Valor-p teste χ^2	<0,001	0,015

Nas próximas duas tabelas, os cálculos foram efetuados apenas dentro dos não aderentes ($n_B=249$ para o Basal e $n_{6m}=187$ para os 6 meses)

Tabela 11. Grau de não adesão no momento Basal e 6 meses

	Basal (n(%))	6 Meses (n(%))
Não aderente esporádico	156 (62,7%)	118 (63,1%)
Não aderente frequente	93 (37,3%)	69 (36,9%)
Valor-p teste χ^2	<0,001	<0,001

Tabela 12. Intencionalidade de não adesão no momento Basal e 6 meses

	Basal (n(%))	6 Meses (n(%))
Não intencional	172 (69,1%)	126 (67,4%)
Intencional	77 (30,9%)	61 (32,6%)
Valor-p teste χ^2	<0,001	<0,001

Tabela 13. Percentagem de resposta Sim na MAT no momento Basal de 6 meses

	Basal (%)	6 Meses (%)
1. Já se esqueceu de tomar os seus medicamentos para a hipertensão?	29,2%	13,2%
2. Já tomou os seus medicamentos fora de horas?	36,3%	32,8%
3. Alguma vez parou de tomar os medicamentos para a pressão arterial por se sentir melhor?	4,5%	4,7%
4. Alguma vez parou de tomar os medicamentos para a pressão arterial por se sentir pior?	0,9%	1,2%
5. Alguma vez aumentou a dose da medicação para a pressão arterial por se sentir pior?	2,6%	3,1%
6. Alguma vez deixou de tomar a medicação para a pressão arterial por deixar terminar a medicação?	8%	5,9%
7. Alguma vez deixou de tomar a medicação para a pressão arterial por outra indicação que não a do médico?	7,3%	2,6%

Tabela 14. Mudanças de adesão no seguimento Basal-6 meses

	n(%)
Sempre Aderente (A_B-A_{6m})	122 (28,8%)
Mudança para pior (A_B-NA_{6m})	53 (12,5%)
Mudança para melhor (NA_B-A_{6m})	115 (27,1%)
Sempre não aderente (NA_B-NA_{6m})	134 (31,6%)
Valor-p teste χ^2	<0,001

Efetuar-se-á agora um estudo semelhante para o seguimento Basal-12meses com **367 indivíduos**. Apesar de este seguimento não ser o mais importante deste trabalho poderá ser interessante comparar as alterações que se encontram quando estamos a trabalhar com um seguimento de 6 meses comparativamente a um seguimento de 12 meses. Por esta razão incluir-se-á este estudo no trabalho.

Tabela 15. Classificação no momento Basal e 12 meses

	Basal (n(%))	12 Meses (n(%))
Aderente	163 (44,4%)	243 (66,2%)
Não Aderente	204 (55,6%)	124 (33,8%)
Valor-p teste χ^2	0,032	<0,001

Nas próximas duas tabelas, os cálculos foram efetuados apenas dentro dos não aderentes ($n_B=204$ para o Basal e $n_{12m}=124$ para os 12 meses)

Tabela 16. Grau de não adesão no momento Basal e 12 meses

	Basal (n(%))	12 Meses (n(%))
Não aderente esporádico	129 (63,2%)	97(78,2%)
Não aderente frequente	75 (36,8%)	27(21,8%)
Valor-p teste χ^2	<0,001	<0,001

Tabela 17. Intencionalidade de não adesão no momento Basal e 12 meses

	Basal (n(%))	12 Meses (n(%))
Não intencional	138 (67,6%)	80 (64,5%)
Intencional	66 (32,4%)	44(35,5%)
Valor-p teste χ^2	<0,001	0,001

Tabela 18. Percentagem de resposta Sim na MAT no momento Basal e 12 meses

	Basal (%)	12 Meses (%)
1. Já se esqueceu de tomar os seus medicamentos para a hipertensão?	26,4%	9,5%
2. Já tomou os seus medicamentos fora de horas?	34,3%	21,5%
3. Alguma vez parou de tomar os medicamentos para a pressão arterial por se sentir melhor?	4,4%	4,6%
4. Alguma vez parou de tomar os medicamentos para a pressão arterial por se sentir pior?	1,1%	1,1%
5. Alguma vez aumentou a dose da medicação para a pressão arterial por se sentir pior?	2,5%	1,6%
6. Alguma vez deixou de tomar a medicação para a pressão arterial por deixar terminar a medicação?	8,7%	5,7%
7. Alguma vez deixou de tomar a medicação para a pressão arterial por outra indicação que não a do médico?	6,8%	4,1%

Tabela 19. Mudanças de adesão no seguimento Basal-12 meses

	n(%)
Sempre Aderente (A_B-A_{12m})	136 (37%)
Mudança para pior (A_B-NA_{12m})	27 (7,4%)
Mudança para melhor (NA_B-A_{12m})	107 (29,2%)
Sempre não aderente (NA_B-NA_{12m})	97 (26,4%)
Valor-p teste χ^2	<0,001

Para cada uma das tabelas apresentadas foi efetuado um teste χ^2 com H_0 : Os dois grupos são iguais. Para todas elas a hipótese nula foi rejeitada pelo que se confirma a diferença entre grupos, para cada instante de tempo.

Não foi aplicado nenhum teste para a comparação de proporções nos dois períodos de tempo, para as características de interesse, pois as amostras não são independentes, não podendo ser aplicado nenhum teste que as considere como tal. Por outro lado também não é possível aplicar nenhum teste para amostras emparelhadas porque há indivíduos que deixaram de participar no estudo durante o tempo de análise.

Através do estudo exploratório dos dados verificamos o seguinte:

Observando as tabelas 10, 11 e 12 para a comparação Basal-6meses, conseguimos verificar que a percentagem de Não Aderentes diminui bastante (quase 15% menos de Não aderentes) ao longo do tempo. Também o pior grau de adesão (Não aderente frequente) diminui quando passamos do Basal para os 6 meses mas com uma diminuição bastante modesta (0,4 pontos percentuais). Já relativamente à intencionalidade, a percentagem de não aderentes intencionais aumenta um pouco (1,7 pontos percentuais) quando questionada aos 6 meses.

Observando a tabela 13 com a percentagem de resposta Sim na MAT para a comparação Basal-6meses (resposta SIM que indica a não adesão à medicação anti-hipertensiva), verifica-se que as duas primeiras respostas, cujo sim indica a não intencionalidade da não adesão, diminuem ao longo do tempo (o que faz sentido pois a não adesão não intencional também diminuiu). Também é nas duas primeiras questões que encontramos as percentagens mais elevadas de respostas Sim, tanto no Basal como nos 6 meses.

Observando a tabela 14, com as mudanças de adesão, observa-se que, entre os 424 indivíduos inquiridos no momento basal e aos 6 meses, 134 (31,6% do total, a

maior percentagem) indica ser sempre não aderente à medicação, enquanto 28,8% indica ser sempre aderente à medicação. Para além disso, 27,1% indicam uma alteração de não aderente no basal para aderente aos 6 meses e 12,5% (a menor percentagem) indicam que mudaram de aderente no momento Basal para não aderente aos 6 meses.

Verificamos, por observação das tabelas 15, 16 e 17 que, quando estudamos o seguimento Basal-12 meses, a percentagem de não aderentes diminui em maior quantidade, chegando aos 21,8 pontos percentuais (contra aproximadamente 15 pontos percentuais no anterior). Também os Não aderentes frequentes têm uma diminuição muito maior relativamente ao seguimento Basal-6meses (15 pontos percentuais contra 0,4 pontos percentuais no anterior). Relativamente aos não aderentes intencionais observa-se um aumento, como no seguimento Basal-6meses, desta vez um pouco maior (3,1 pontos percentuais contra 1,7 pontos percentuais).

A tabela 18 com a percentagem de respostas 'sim' no seguimento Basal-12meses é bastante semelhante à tabela do seguimento Basal-6meses: as maiores percentagens de respostas sim estão nas duas primeiras questões, com valores mais baixos aos 12 meses, em particular na segunda pergunta que desce 12,8% (contra 3,5% no seguimento Basal-6meses). A questão número 5 apresenta uma ligeira subida ao contrário do que acontecia no seguimento Basal-6meses mas com percentagens muito pequenas.

Observando a tabela 19, verificamos que 37% dos aderentes se mantiveram aderentes ao fim de 12 meses (um pouco mais que no seguimento Basal-6 meses com 28,8%). A percentagem de Sempre não aderentes é um pouco menor no seguimento de 12 meses (26,4% contra 31,6% no seguimento de 6 meses). A mudança Não aderente para aderente aumenta ligeiramente para o seguimento de 12 meses, de 27,1% no seguimento de 6 meses contra 29,2% no seguimento de 12 meses.

5.2 Regressão logística binária

Apesar de este trabalho ser focado na variável ‘Mudanças’, será interessante verificar o que acontece com a variável ‘Classificação’ no momento Basal. Nesta secção proceder-se-á ao ajustamento de regressões logísticas binárias nos dados em estudo, com estas duas variáveis como variável dependente. Todo o processo de ajustamento dos modelos será realizado com base no software R, versão 3.0.1. A seleção das variáveis significativas para descrever os vários fenómenos e causas, será efetuada recorrendo a duas técnicas distintas: “método empírico” e método stepwise.

O “método empírico” funciona da seguinte forma:

Efetuar-se-ão regressões logísticas simples de todas as variáveis com a variável dependente Classificação ou Mudança de adesão. As variáveis cujo valor p seja menor que 0,15 entrarão no modelo. Seguidamente efetuar-se-á a regressão logística múltipla com a variável dependente Classificação ou Mudança de adesão e as variáveis independentes escolhidas para entrar.

5.2.1 Regressão logística: variável dependente Classificação (Basal)

A variável “Classificação” indica se um doente é classificado como “Aderente” ou “Não Aderente”, com 0-Aderente e 1-Não aderente. Na secção 5.1 foi efetuado um estudo exploratório para esta mesma variável. O método empírico começará com um estudo univariado presente na tabela 20.

Com a utilização do método empírico obtiveram-se os resultados apresentados na tabela 20.

Tabela 20. Dados relativos às regressões logísticas simples com variável dependente 'Classificação'. As variáveis a entrar no modelo têm valor $p \leq 0,15$

Variável	Aderentes n=115	Não aderentes n=170	Valor p	OR	IC OR (95%)	β
Idade, média \pm dp	62,66 \pm 9,82	58,80 \pm 10,43	0,002	0,96	[0,94;0,99]	-0,037
Género, n (%)						
Feminino	65 (56,5%)	84 (49,4%)	0,239	Ref	[0,83;2,14]	0,286
Masculino	50 (43,5%)	86 (50,6%)		1,33		
Ocupação, n (%)						
Trabalhador/Estudante	39 (33,9%)	73(42,9%)	0,719	Ref	[0,54;2,47]	0,140
Desempregado	13 (11,3%)	28(16,5%)		1,15		
Reformado/Doente	63 (54,8%)	69(40,6%)		0,042		
Estado Civil, n (%)						
Solteiro	5 (4,3%)	23 (13,5%)	0,025	Ref	[0,12;0,87]	-1,152
Casado/União Facto	86(74,8%)	125 (73,5%)		0,32		
Divorciado/Separado	12(10,4%)	12(7,1%)		0,017		
Viúvo	12(10,4%)	10(5,9%)		0,22		
Viúvo	12(10,4%)	10(5,9%)	0,009	0,18	[0,05;0,65]	-1,708
Anos de escolaridade completos, média \pm dp	6,70 \pm 4,15	6,96 \pm 3,86	0,574	1,02	[0,96;1,08]	0,017
Etnia, n (%)						
Etnia Branca	76 (66,1%)	88(51,8%)	0,017	Ref	[1,11;2,96]	0,597
Etnia Negra	39(33,9%)	82(48,2%)		1,82		
Vive com quantas pessoas, média \pm dp	1,99 \pm 1,42	2,18 \pm 1,59	0,300	1,09	[0,93;1,28]	0,084
Dificuldade financeira para comprar medicação, n (%)						
Sem dificuldade	55 (47,8%)	65 (38,2%)	0,072	Ref	[0,95;2,91]	0,510
Alguma dificuldade	32 (27,8%)	63 (37,1%)		1,67		
Muita dificuldade	28 (24,3%)	42 (24,7%)		0,435	[0,70;2,31]	0,238
Idade de diagnóstico de HTA, média \pm dp	48,55 \pm 11,74	45,98 \pm 12,13	0,078	0,98	[0,96;1]	-0,018
Número de medicamentos, média \pm dp	4,97 \pm 2,93	4,59 \pm 2,66	0,250	0,95	[0,87;1,04]	-0,050
Número medicamentos para HTA, Média \pm dp	1,71 \pm 0,97	1,81 \pm 0,90	0,380	1,12	[0,87;1,46]	0,117
IMC, n (%)						
Baixo peso/normal	15 (13%)	21(12,4%)	0,784	Ref	[0,53;2,39]	0,107
Excesso de peso	43 (37,4%)	67(39,4%)		1,11		
Obesidade	57 (49,6%)	82(48,2%)		0,943	[0,49;2,16]	0,027
Tem diabetes (n%)						
não	81 (70,4%)	126(74,1%)	0,494	Ref	[0,49;1,41]	-0,184
sim	34 (29,6%)	44(25,9%)		0,83		
Tem colesterol (n%)						
sim	62(53,9%)	81(47,6%)	0,300	Ref	[0,80;2,07]	0,251
não	53(46,1%)	89 (52,4%)		1,29		
Risco cardiovascular (n%)						
sem	29(25,2%)	45(26,5%)	0,813	Ref	[0,55;1,61]	-0,065
com	86(74,8%)	125(73,5%)		0,94		
Considera HTA controlada (n%)						
sim	96(83,5%)	125(73,5%)	0,050	Ref	[1,00;3,31]	0,598
não	19(16,5%)	45(26,5%)		1,82		

Tabela 20. Dados relativos às regressões logísticas simples com variável dependente 'Classificação'. As variáveis a entrar no modelo têm valor $p \leq 0,15$ (cont.)

Variável	Aderentes n=115	Não aderentes n=170	Valor p	OR	IC OR (95%)	β
Tem ajuda para controlar HTA (%)						
sim	10(8,7%)	30(17,6%)		Ref		
não	105(91,3%)	140(82,4%)	0,036	0,44	[0,21;0,95]	-0,810
Frequência da medição HTA (%)						
Medição regular	83(72,2%)	128(75,3%)		Ref		
Medição pouco regular	29(25,2%)	35(20,6%)	0,395	0,78	[0,59;1,02]	-0,25
Outras	3(2,6%)	7(4,1%)	0,557	0,88	[0,38;1,17]	-0,41
Recorreu ao médico de família no último ano (%)						
sim	114 (99,1%)	167 (98,2%)		Ref		
não	1 (0,9%)	3 (1,8%)	0,537	2,05	[0,21;19,90]	0,717
Recorreu ao médico especialista no último ano (%)						
sim	8 (7%)	17 (10%)		Ref		
não	107 (93%)	153 (90%)	0,375	0,67	[0,28;1,62]	-0,396
Recorreu ao médico privado no último ano (%)						
sim	7 (6,1%)	16 (9,4%)		Ref		
não	108 (93,9%)	154 (90,6%)	0,316	0,62	[0,25;1,57]	-0,472
Recorreu ao serviço de urgência no último ano (%)						
sim	10 (8,7%)	19 (11,2%)		Ref		
não	105 (91,3%)	151 (88,8%)	0,498	0,76	[0,34;1,69]	-0,279
Recorreu ao atendimento no centro de saúde no último ano (%)						
sim	9 (7,8%)	10 (5,9%)		Ref		
não	106 (92,2%)	160 (94,1%)	0,520	1,36	[0,53;3,45]	0,306
Recorreu ao farmacêutico no último ano (%)						
sim	30 (26,1%)	62 (36,5%)		Ref		
não	85 (73,9%)	108 (63,5%)	0,067	0,61	[0,37;1,03]	-0,487
Foi seguido regularmente na farmácia no último ano (%)						
sim	20 (17,4%)	25 (14,7%)		Ref		
não	95 (82,6%)	145 (85,3%)	0,542	1,22	[0,64;2,32]	0,200
Recorreu ao enfermeiro no último ano (%)						
sim	7 (6,1%)	16 (9,4%)		Ref		
não	108 (93,9%)	154 (90,6%)	0,316	0,62	[0,25;1,57]	- 0,472
Controlo da TA 140/90 (%)						
controlado	53 (46,1%)	79 (46,5%)		Ref		
não controlado	62 (53,9%)	91 (53,5%)	0,949	0,98	[0,61;1,58]	-0,015
Média pressão arterial sistólica, média \pm dp	141,51 \pm 19,78	139,43 \pm 19,52	0,38	1	[0,98;1,01]	-0,005
Média pressão arterial diastólica, média \pm dp	84,58 \pm 11,19	86,09 \pm 11,40	0,268	1,01	[0,99;1,03]	0,012

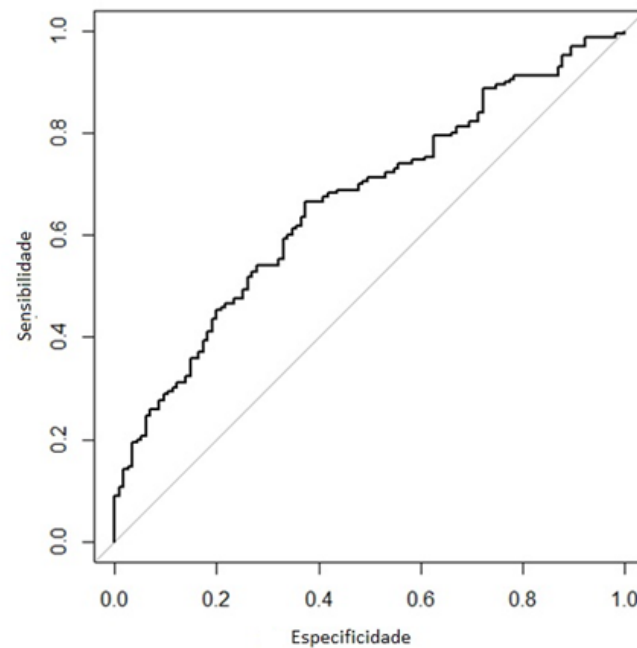
Observando a tabela verifica-se que as variáveis a incluir no modelo de regressão logística múltipla são: Idade, Ocupação, Estado Civil, Etnia, Dificuldade financeira na compra de medicação, Idade de diagnóstico de HTA, Considera HTA controlada, Tem ajuda para controlar HTA, Recorreu ao farmacêutico no último ano. Com base nestas análises univariadas, procedeu-se ao ajustamento do modelo de regressão logística binária múltipla. Os resultados encontram-se na tabela 21.

Tabela 21. Resultados da regressão logística múltipla com variável dependente Classificação utilizando o método empírico.

Variáveis	β	Valor p	OR
Idade	-0,020	0,283	0,98
Ocupação			
Trabalhador/Estudante			Ref
Desempregado	0,257	0,528	1,29
Reformado/Doente	-0,041	0,913	0,96
Estado Civil			
Solteiro			Ref
Casado/União de Facto	-0,991	0,076	0,37
Divorciado/Separado	-1,217	0,072	0,30
Viúvo	-1,367	0,050	0,25
Etnia			
Branca			Ref
Negra	0,142	0,667	1,15
Dificuldade na compra de medicação			
Sem dificuldade			Ref
Alguma dificuldade	0,299	0,326	1,35
Muita dificuldade	-0,147	0,685	0,86
Idade de diagnóstico de HTA	-0,005	0,727	1
Considera HTA controlada			
Sim			Ref
Não	0,430	0,191	1,54
Tem ajuda para controlar HTA?			
Sim			Ref
Não	-0,762	0,061	0,47
Recorreu ao farmacêutico no último ano			
Sim			Ref
Não	-0,270	0,383	0,76
Teste de Hosmer e Lemeshow	Valor P= 0,679		

Analisando a tabela 21 verifica-se que as variáveis significativas para o modelo são: Estado Civil (Casado, Divorciado e Viúvo) e Tem ajuda para controlar HTA (valor $p < 0,10$).

Figura 2. Gráfico de curva ROC para o modelo de regressão logística múltipla com variável dependente Classificação utilizando o método empírico.



A área abaixo da curva ROC é de 0,6644.

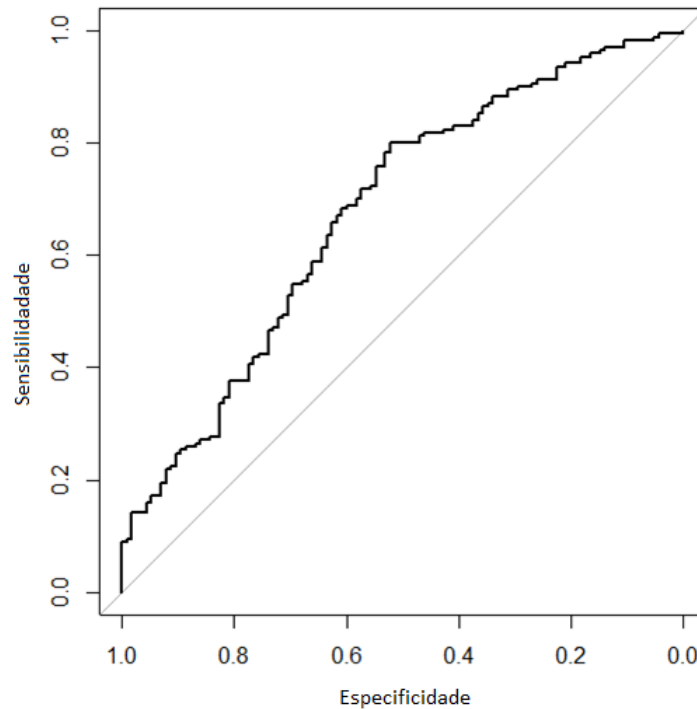
Utilizando o método Stepwise, com todas as variáveis apresentadas na tabela 20, obteve-se o modelo de regressão logística apresentado na tabela 22.

Tabela 22. Resultados da regressão logística múltipla com variável dependente Classificação utilizando o método Stepwise.

Variáveis	β	Valor p	OR
Idade	-0,035	0,017	0,97
Género			
Feminino			Ref
Masculino	0,710	0,010	2,03
Etnia			
Branca			Ref
Negra	0,702	0,023	2,02
Vive com quantas pessoas	-0,147	0,154	0,86
Considera HTA controlada			
Sim			Ref
Não	0,807	0,025	2,24
Tem ajuda para controlar HTA			
Sim			Ref
Não	-0,842	0,037	0,43
Media pressão arterial sistólica	-0,012	0,103	0,99
Teste de Hosmer e Lemeshow	Valor P= 0,578		

Utilizando o método Stepwise encontram-se algumas diferenças. As variáveis significativas para o modelo são: Idade, Género, Etnia, Considera HTA controlada e Tem ajuda para controlar HTA. (valor p <0,10)

Figura 3. Gráfico de curva ROC para o modelo de regressão logística múltipla com variável dependente Classificação utilizando o método Stepwise.



A área abaixo da curva ROC é de 0,6764.

Utilizando o modelo de regressão logística, em que a variável resposta é a “Classificação”, observando a tabela 20 que nos indica as relações bivariadas entre cada uma das covariáveis e a variável resposta podemos verificar para as variáveis significativas ($p \leq 0,15$) o seguinte: A etnia negra tem mais 82% de hipótese de ser “não aderente” do que a etnia branca; pessoas com dificuldades financeiras para comprar os medicamentos têm mais 67% de hipótese de ser não aderentes do que pessoas sem dificuldade financeira (OR= 1,67), indivíduos que não consideram a HTA controlada têm mais 82% de hipótese de serem “não aderentes” à medicação (OR= 1,82); quanto mais velho um indivíduo é menos hipótese tem de ser não aderente à medicação (OR=0,96). Verifica-se ainda que pessoas referenciadas como Reformados/Doentes /Permanentemente Incapacitados têm menos 41% de hipótese de ser “não aderentes” do que os classificados como Trabalhadores/Estudantes/Dona de casa (OR= 0,59); Todos os estados civis são menos “não aderentes” comparando com o Solteiro: Casado com menos 68% de hipótese de ser “não aderente” (OR=0,32), Divorciado com menos 78% de hipótese (OR=0,22) e Viúvos com menos 82% de hipótese de ser não aderentes (OR=0,18); quanto mais tarde for diagnosticada a hipertensão arterial no doente, maior a probabilidade de ser aderente à medicação (OR=0,98); indivíduos que não têm ajuda para controlar a Hipertensão têm menos 56% de hipótese de serem “não aderentes” do que os que têm ajuda; indivíduos que não foram ao farmacêutico no último ano têm menos hipótese de ser “não aderentes” (menos 39% de hipótese) que indivíduos que recorreram.

Com estas variáveis construímos o modelo logístico múltiplo. As variáveis escolhidas para fazerem parte do modelo foram Idade, Ocupação, Estado Civil, Etnia, Dificuldade financeira para comprar medicamentos, Idade de diagnóstico de HTA, considera HTA controlada, Tem ajuda para controlar HTA, Recorreu ao farmacêutico no último ano.

A tabela 21 indica as variáveis associadas com a não adesão à medicação com HTA através de uma regressão logística múltipla. As variáveis que se relacionam de forma importante com a não adesão são o **Estado Civil e Tem ajuda para controlar HTA**, sendo que novamente todos os estados civis parecem ser um fator propiciador da adesão quando comparados com o estado Solteiro (Casado com menos 63% de hipótese de ser não aderente, Divorciado com menos 70% de hipótese e Viúvo com menos 75% de hipótese de ser não aderente). Estes resultados poderão ter um problema pois existem apenas 5 solteiros na amostra utilizada. Também o facto de não ter ajuda para controlar a hipertensão parece ser um fator propiciador da adesão (indivíduos sem ajuda para controlar a hipertensão com mais 53% de hipótese de ser aderentes comparando com os indivíduos que têm ajuda). Para esta regressão foi efetuado o teste de ajustamento através da curva ROC com $AUC = 0,6644$ que indica que o modelo está ajustado.

Utilizando o método Stepwise de escolha das variáveis a entrar no modelo de regressão logística múltipla os resultados foram bastante diferentes, não só ao nível das variáveis que fazem parte do modelo (Idade, Género, Etnia, Vive com quantas pessoas, Pensa que tem HTA controlado, Tem ajuda para controlar HTA, Media Pressão arterial sistólica) mas também ao nível dos resultados (tabela 22.). As variáveis importantes para o modelo são: **Idade, Género, Etnia, considera HTA controlada e Tem ajuda para controlar HTA**. Para a etnia negra há o dobro das hipóteses de ser “não aderentes” comparando com a etnia branca assim como indivíduos que consideram a hipertensão não controlada têm mais do dobro das hipóteses de ser “não aderentes”; homens têm o dobro das hipóteses de serem “não aderente” comparando com as mulheres. A variável idade parece ser um fator propiciador da adesão, quanto mais velha for a pessoa maior é a tendência para ser “aderente” à medicação. O facto de não ter ajuda para controlar a hipertensão parece ser um fator que incita à adesão (indivíduos sem ajuda para controlar a hipertensão com mais 57% de hipótese de serem “aderentes” comparando com os indivíduos que têm ajuda). Para esta regressão foi efetuado o teste de ajustamento através da curva ROC com $AUC=0,6764$ que indica que o modelo está ajustado.

5.2.2 Regressão logística: variável dependente Mudança de adesão (com duas categorias)

A variável “Mudança” indica se um doente é classificado como 0 - “Mudança positiva” ou 1 - “Mudança negativa”. Na categoria “Mudança positiva” foram incluídos os indivíduos que se mantiveram “aderentes” no momento Basal e 6 meses e os indivíduos que mudaram de “não aderentes” no momento basal para “aderentes” aos 6 meses; já na categoria “Mudança negativa” inserem-se os indivíduos que se mantiveram “não aderentes” no momento basal e 6 meses e os indivíduos que mudaram de “aderentes” no momento basal para “não aderentes” aos 6 meses. Na secção 5.1 foi efetuado um estudo exploratório para esta mesma variável. O método empírico começará com um estudo univariado presente na tabela 23.

Utilizando o “método empírico” obteve-se a tabela 23.

Tabela 23. Dados relativos às regressões logísticas simples com variável dependente Mudança de adesão. As variáveis a entrar no modelo têm valor $p \leq 0,15$

Variável	Aderentes n=162	Não aderentes n=123	Valor p	OR	IC OR (95%)	β
Idade, média \pm dp	62,23 \pm 9,57	57,89 \pm 10,84	<0,001	0,96	[0,94;0,98]	
Género, n (%)						
Feminino	79 (48,8%)	70 (56,9%)	0,173	Ref		
Masculino	83 (51,2%)	53 (43,1%)		0,72	[0,45;1,15]	-0,328
Ocupação, n (%)						
Trabalhador/Estudante	46 (28,4%)	66(53,7%)		Ref		
Desempregado	25 (15,4%)	16(13%)	0,031	0,45	[0,21;0,93]	-0,807
Reformado/Doente	91 (56,2%)	41(33,3%)	<0,001	0,31	[0,19;0,53]	-1,158
Estado Civil, n (%)						
Solteiro	10 (6,2%)	18 (14,6%)		Ref		
Casado/União Facto	125 (77,2%)	86 (69,9%)	0,022	0,38	[0,17;0,87]	-0,962
Divorciado/Separado	11(6,8%)	13 (10,6%)	0,459	0,66	[0,22;2,00]	-0,421
Viúvo	16(9,9%)	6 (4,9%)	0,011	0,21	[0,06;0,70]	-1,569
Anos de escolaridade completos, média \pm dp	7,2 \pm 4,05	6,41 \pm 3,83	0,097	0,95	[0,89;1,01]	-0,051
Etnia, n (%)						
Etnia Branca	119 (73,5%)	45 (36,6%)		Ref		
Etnia Negra	43 (26,5%)	78 (63,4%)	<0,001	4,80	[2,89;7,96]	1,568
Vive com quantas pessoas, média \pm dp	1,69 \pm 1,19	2,66 \pm 1,74	<0,001	1,60	[1,33;1,91]	0,467
Dificuldade financeira para comprar medicação, n (%)						
Sem dificuldade	74 (45,7%)	46 (37,4%)		Ref		
Alguma dificuldade	57 (35,2%)	38 (30,9%)	0,804	1,07	[0,62;1,86]	0,070
Muita dificuldade	31 (19,1%)	39 (31,7%)	0,021	2,02	[1,11;3,68]	0,705
Idade de diagnóstico de HTA, média \pm dp	48,57 \pm 12,25	44,97 \pm 11,44	0,013	0,97	[0,96;0,99]	-0,026
Número de medicamentos, média \pm dp	4,88 \pm 2,91	4,57 \pm 2,58	0,354	0,96	[0,88;1,05]	-0,041
Número medicamentos para HTA, Média \pm dp	1,73 \pm 0,94	1,82 \pm 0,92	0,437	1,11	[0,86;1,42]	0,100
IMC, n (%)						
Baixo peso/normal	23 (14,2%)	13 (10,46%)		Ref		
Excesso de peso	58 (35,8%)	52 (42,3%)	0,244	1,59	[0,73;3,45]	0,461
Obesidade	81 (50%)	58 (47,2%)	0,541	1,27	[0,59;2,71]	0,237
Tem diabetes (n%)						
não	122 (75,3%)	85 (69,1%)		Ref		
sim	40 (24,7%)	38 (30,9%)	0,246	0,73	[0,81;2,30]	-0,310
Tem colesterol (n%)						
sim	82 (50,6%)	61 (49,6%)		Ref		
não	80 (49,4%)	62 (50,4%)	0,864	1,04	[0,65;1,66]	0,041
Risco cardiovascular (n%)						
sem	43 (26,5%)	31 (25,2%)		Ref		
com	119 (73,5%)	92 (74,8%)	0,798	0,93	[0,63;1,83]	-0,070
Considera HTA controlada (n%)						
sim	135 (83,3%)	86 (69,9%)		Ref		
não	27 (16,7%)	37 (30,15%)	0,008	2,15	[1,22;3,79]	0,766

Tabela 23. Dados relativos às regressões logísticas simples com variável dependente Mudança de adesão. As variáveis a entrar no modelo têm valor $p \leq 0,15$ (cont.).

Variável	Aderentes n=162	Não aderentes n=123	Valor p	OR	IC OR (95%)	β
Tem ajuda para controlar HTA (%)						
sim	18 (11,1%)	22 (17,9%)		Ref		
não	144(88,9%)	101(82,1%)	0,106	0,57	[0,29;1,12]	-0,555
Frequência da medição HTA (%)						
Medição regular	119 (73,5%)	92 (74,8%)		Ref		
Medição pouco regular	37 (22,8%)	27 (22%)	0,842	0,94	[0,54;1,66]	-0,058
Outras	6 (3,7%)	4 (3,3%)	0,823	0,86	[0,24;3,15]	-0,148
Recorreu ao médico de família no último ano (%)						
sim	160 (98,8%)	121 (98,4%)		Ref		
não	2 (1,2%)	2 (1,6%)	0,782	1,32	[0,18;9,52]	0,279
Recorreu ao médico especialista no último ano (%)						
sim	13 (8%)	12 (9,8%)		Ref		
não	149 (92%)	111 (90,2%)	0,609	0,81	[0,35;1,84]	-0,214
Recorreu ao médico privado no último ano (%)						
sim	13 (8%)	10 (8,1%)		Ref		
não	149 (92%)	113 (91,9%)	0,974	0,99	[0,42;2,33]	-0,014
Recorreu ao serviço de urgência no último ano (%)						
sim	12 (7,4%)	17(13,8%)		Ref		
não	150 (92,6)	106 (86,2%)	0,08	0,50	[0,23;1,09]	-0,696
Recorreu ao atendimento no centro de saúde no último ano (%)						
sim	10 (6,2%)	9 (7,3%)		Ref		
não	152 (93,8%)	114 (92,7%)	0,702	1,20	[0,33;2,12]	0,182
Recorreu ao farmacêutico no último ano (%)						
sim	34 (21%)	58 (47,2%)		Ref		
não	128 (79%)	65 (52,8%)	<0,001	0,30	[0,18;0,50]	-1,212
Foi seguido regularmente na farmácia no último ano (%)						
sim	24 (14,8%)	21 (17,1%)		Ref		
não	138 (85,2%)	102 (82,9%)	0,605	0,84	[0,45;1,60]	-0,169
Recorreu ao enfermeiro no último ano (%)						
sim	5 (3,1%)	18 (14,6%)		Ref		
não	157 (96,9%)	105 (85,4%)	0,001	0,19	[0,07;0,52]	- 1,683
Controlo da TA 140/90 (%)						
controlado	83 (51,2%)	49 (39,8%)		Ref		
não controlado	79 (48,8%)	74 (60,2%)	0,057	1,59	[0,99;2,55]	0,462
Média pressão arterial sistólica, média \pm dp	139,25 \pm 19,03	141,62 \pm 20,37	0,314	0,99	[0,99;1,02]	-0,006
Média pressão arterial diastólica, média \pm dp	83,23 \pm 11,24	88,45 \pm 10,77	<0,001	1,04	[1,02;1,07]	0,043

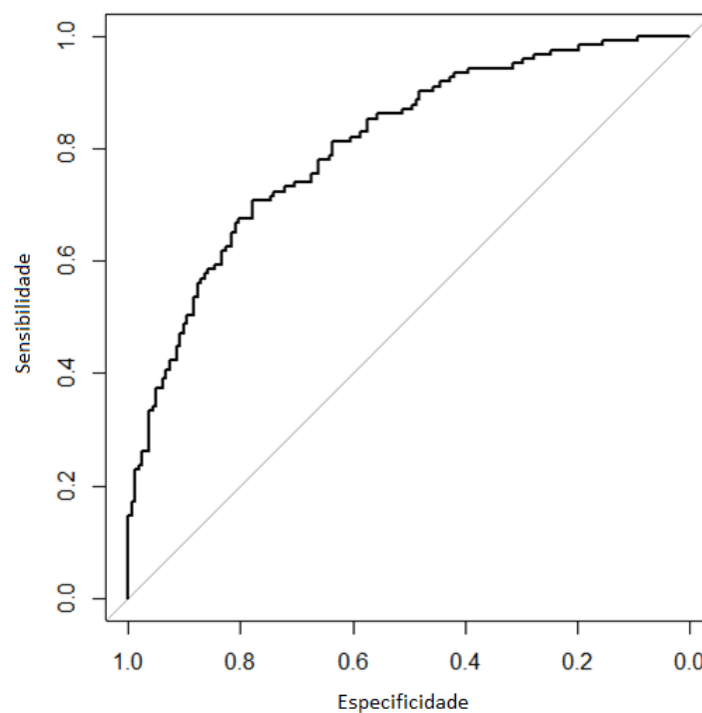
Observando a tabela 23 verifica-se que as variáveis a incluir no modelo de regressão logística multivariado são: Idade; Ocupação; Estado Civil; Anos de escolaridade completos; Etnia; vive com quantas pessoas; Dificuldade financeira para comprar medicamentos; Idade de diagnóstico de HTA; considera HTA controlada; tem ajuda para controlar HTA; Recorreu ao serviço de urgência no último ano; Recorreu ao farmacêutico no último ano; Recorreu ao enfermeiro no último ano; Controlo da TA e Média da pressão arterial diastólica. O modelo logístico múltiplo obtido é apresentado na tabela 24.

Tabela 24. Resultados da regressão logística múltipla com variável dependente Mudança de adesão utilizando o método empírico.

Variáveis	β	Valor p	OR
Idade	0,007	0,749	1,01
Ocupação			
Trabalhador/Estudante			Ref
Desempregado	-0,801	0,066	0,45
Reformado/Doente	-0,538	0,219	0,58
Estado Civil			
Solteiro			Ref
Casado/União	0,364	0,517	1,44
Divorciado/Separado	1,757	0,019	5,80
Viúvo	0,007	0,992	1,01
Anos escolaridade completos	-0,059	0,141	0,94
Etnia			
Branca			Ref
Negra	1,185	0,002	3,27
Vive com quantas pessoas	0,216	0,086	1,24
Dificuldade financeira na compra da medicação			
Sem dificuldade			Ref
Alguma dificuldade	-0,841	0,023	0,43
Muita dificuldade	-0,455	0,285	0,63
Idade diagnóstico de HTA	-0,005	0,727	1
Considera HTA controlada			
Sim			Ref
Não	0,219	0,575	1,24
Precisa de ajuda para controlar a HTA			
Sim			Ref
Não	-0,690	0,091	0,51
Recorreu ao serviço de urgência no último ano			
Sim			Ref
Não	-0,186	0,719	0,83
Recorreu ao farmacêutico no último ano			
Sim			Ref
Não	-0,924	0,006	0,40
Recorreu ao enfermeiro no último ano			
Sim			Ref
Não	-1,713	0,005	0,18
Controlo da TA			
Controlado			Ref
Não controlado	-0,236	0,552	0,79
Média da pressão arterial diastólica	0,037	0,051	1,04
Teste de Hosmer e Lemeshow	Valor P= 0,987		

As variáveis significativas para o modelo são: Ocupação (Desempregado), Estado Civil (Divorciado/Separado), Etnia, Vive com quantas pessoas, Dificuldade financeira na compra da medicação, Tem ajuda para controlar a HTA, Recorreu ao farmacêutico no último ano, Recorreu ao enfermeiro no último ano e Média da pressão arterial diastólica. (valor $p < 0,10$).

Figura 4. Gráfico de curva ROC para o modelo de regressão logística múltipla com variável dependente Mudança de adesão utilizando o método empírico.



A área abaixo da curva ROC é de 0,8066.

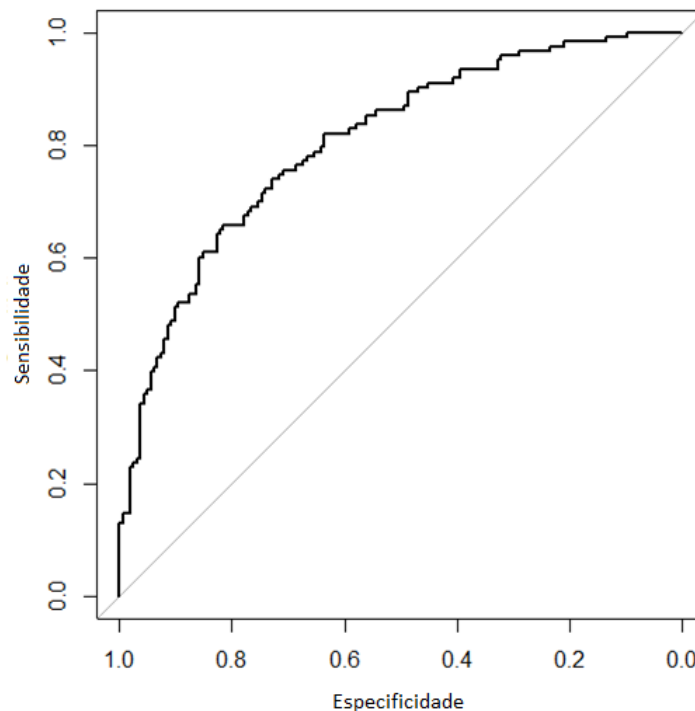
Utilizando o método Stepwise obtenho o seguinte modelo de regressão logística:

Tabela 25. Resultados da regressão logística múltipla com variável dependente Mudança de adesão utilizando o método Stepwise.

Variáveis	β	Valor p	OR
Ocupação			Ref
Trabalhador/Estudante			0,45
Desempregado	-0,807	0,063	0,60
Reformado/Doente	-0,509	0,147	
Estado Civil			Ref
Solteiro			1,52
Casado/União	0,418	0,453	6,41
Divorciado/Separado	1,858	0,012	1,10
Viúvo	0,091	0,902	
Anos escolaridade completos	-0,064	0,096	0,94
Etnia			Ref
Branca			3,31
Negra	1,198	0,001	
Vive com quantas pessoas	0,211	0,073	1,23
Dificuldade financeira para comprar medicamentos			Ref
Sem dificuldade			0,45
Alguma dificuldade	-0,806	0,025	0,68
Muita dificuldade	-0,392	0,347	
Tem ajuda para controlar HTA			Ref
Sim			0,50
Não	-0,689	0,090	
Recorreu ao farmacêutico no último ano			Ref
Sim			0,40
Não	-0,921	0,004	
Recorreu ao enfermeiro no último ano			Ref
Sim			0,18
Não	-1,726	0,005	
Média da pressão artéria diastólica	0,033	0,019	1,03
Teste de Hosmer e Lemeshow	Valor P= 0,990		

Neste caso o método Stepwise obteve um resultado muito semelhante ao método empírico, com apenas uma variável a mais. As variáveis significativas no modelo são: Ocupação (Desempregado), Estado Civil (Divorciado/Separado), Anos de escolaridade completos, Etnia, Vive com quantas pessoas, Dificuldade financeira para comprar medicação (Alguma dificuldade), Tem ajuda para controlar HTA, Recorreu ao farmacêutico no último ano, Recorreu ao enfermeiro no último ano e Média da pressão artéria diastólica (valor $p < 0,10$).

Figura 5. Gráfico de curva ROC para o modelo de regressão logística múltipla com variável dependente Mudança de adesão utilizando o método Stepwise.



A área abaixo da curva ROC é de 0,8061.

Com a regressão logística com variável resposta “Mudança”, com 2 categorias, observando a tabela 23 que nos indica as relações bivariadas entre cada uma das covariáveis e a variável resposta podemos verificar para as variáveis significativas ($p \leq 0,15$) o seguinte: a etnia negra tem quase 5 vezes mais hipóteses de ter uma “mudança negativa” que a etnia branca; viver com várias pessoas na mesma casa indica maior hipótese de ter uma “mudança negativa”; pessoas com muita dificuldade para comprar os medicamentos têm 2 vezes mais hipótese de ter uma “mudança negativa” que pessoas sem dificuldade financeira para comprar a medicação; pessoas que consideram a sua HTA não controlada têm o dobro das hipóteses de ter uma “mudança negativa” comparativamente aos que a consideram controlada; indivíduos com a pressão arterial não controlada têm mais 59% de hipótese de ter uma “mudança negativa” que indivíduos com a pressão arterial controlada. Quanto maior a Média da pressão arterial diastólica maior as hipóteses de ter uma “mudança negativa”. Verifica-se ainda que quanto mais idade um indivíduo tiver mais hipóteses de ter uma “mudança positiva”; Indivíduos desempregados têm 55% mais de hipóteses de ter uma “mudança positiva” e indivíduos Reformados/Doentes/permanentemente incapacitados mais 69% de apresentar uma “mudança positiva” que os indivíduos Trabalhadores/ Estudantes/ Donas de Casa; Casados têm mais 62% de hipótese de ter uma “mudança positiva”, assim como Viúvos têm mais 79% de ter a mesma mudança

comparativamente a indivíduos solteiros; quanto maior o nível de escolaridade mais hipótese existe de ter uma “mudança positiva”; quanto maior a idade em que é descoberta a hipertensão arterial menor são as hipóteses de ter uma “mudança negativa”; pessoas sem ajuda para controlar a HTA têm menos 43% de hipótese de ter uma “mudança negativa” comparativamente com pessoas com ajuda; indivíduos que não recorreram ao serviço de urgência no último ano têm 50% mais probabilidade de ter uma “mudança positiva” que indivíduos que recorreram às urgências; indivíduos que não recorreram ao farmacêutico no último ano têm 70% mais hipótese de ter uma “mudança positiva” que indivíduos que recorreram ao farmacêutico; indivíduos que não recorreram ao enfermeiro no último ano têm mais 81% de hipótese de ter uma “mudança positiva” que indivíduos que não recorreram ao enfermeiro.

As variáveis a incluir no modelo de regressão logística múltipla são: Idade, Ocupação, Estado Civil, Anos de escolaridade completos, Etnia, vive com quantas pessoas, Dificuldade financeira para comprar medicamentos, Idade de diagnóstico de HTA, considera HTA controlada, tem ajuda para controlar HTA, Recorreu ao serviço de urgência no último ano, Recorreu ao farmacêutico no último ano, Recorreu ao enfermeiro no último ano, Controlo da TA e Média da pressão arterial diastólica.

Observando a tabela 24. Verificamos que as variáveis importantes para o modelo são: **Ocupação, Estado Civil, Etnia, vive com quantas pessoas, Dificuldade financeira para comprar medicamentos, tem ajuda para controlar HTA, Recorreu ao farmacêutico no último ano, Recorreu ao enfermeiro no último ano e Média da pressão arterial diastólica** (valor $p < 0,10$), sendo que as pessoas divorciadas têm 5.8 vezes mais hipótese de ter uma “mudança negativa” quando comparadas com as solteiras; a etnia negra tem três vezes mais hipótese de ter uma “mudança negativa” que a etnia branca; um maior número de pessoas na mesma casa aumenta o risco de uma “mudança negativa” ($OR = 1.24$) e o mesmo acontece quanto maior for a média da pressão arterial diastólica. Já a variável Ocupação (comparação Desempregado – trabalhador/estudante/dona de casa), dificuldade financeira (comparação alguma dificuldade – nenhuma dificuldade), ajuda a controlar a HTA, recorreu ao farmacêutico no último ano e recorreu ao enfermeiro no último ano parecem conduzir a uma “mudança positiva”.

Utilizando o método Stepwise de escolha das variáveis a entrar no modelo de regressão logística múltipla os resultados foram: Ocupação, Estado Civil, anos de escolaridade completos, Etnia, vive com quantas pessoas, Dificuldade financeira para comprar medicamentos, tem ajuda para controlar HTA, Recorreu ao farmacêutico no último ano, Recorreu ao enfermeiro no último ano e Média da pressão arterial diastólica. Apesar de as variáveis a entrar no modelo serem ligeiramente diferentes das utilizadas no modelo empírico (menos variáveis) os resultados finais foram muito semelhantes. As variáveis importantes para o modelo logístico multivariado são

Ocupação, Estado Civil, anos de escolaridade completos, Etnia, vive com quantas pessoas, Dificuldade financeira para comprar medicamentos, tem ajuda para controlar HTA, Recorreu ao farmacêutico no último ano, Recorreu ao enfermeiro no último ano e Média da pressão arterial diastólica, todas com valores muito próximos de odds ratio do “modelo empírico”, logo com a mesma interpretação. As exceções são a variável anos de escolaridade completos que indica que um maior número de anos na escola é fator propiciador da mudança positiva e o Estado Civil que apesar de ser em tudo semelhante ao modelo empírico passou de cinco vezes superior para mais de seis.

5.3 Regressão logística multinomial com variável dependente Mudança de adesão (com três categorias)

A regressão multinomial será realizada com a variável dependente Mudança de adesão com três categorias (já indicadas na tabela 9). Na regressão logística (duas categorias) são comparadas as duas categorias existentes. No caso da regressão multinomial, existe uma categoria para comparação com as restantes categorias. A categoria escolhida para comparação foi a categoria 0-Sempre Aderente. Assim, obteremos duas tabelas, uma com a comparação 0-Sempre Aderente vs 1-Melhor comportamento de adesão e uma segunda com a comparação 0-Sempre Aderente vs 2- Mudança/Manutenção de não adesão.

Assim, com base na análise univariada obteve-se a tabela 26, com a comparação 0-Sempre Aderente vs 1-Melhor comportamento de adesão:

Tabela 26. Dados relativos às regressões logísticas multinomiais simples com variável dependente Mudança de adesão: comparação 0-Sempre Aderente vs 1-Melhor comportamento. As variáveis a entrar no modelo têm valor $p \leq 0,15$.

Variável	Aderentes n=79	Não aderentes n=83	Valor p	OR	IC OR (95%)	β
Idade, média \pm dp	63,35 \pm 9,64	61,17 \pm 9,44	0,170	0,98	[0,95;1,01]	-0,022
Género, n (%)						
Feminino	41 (51,9%)	38 (45,8%)		Ref		
Masculino	38 (48,1%)	45 (54,2%)	0,437	1,28	[0,69;2,37]	0,245
Ocupação, n (%)						
Trabalhador/Estudante	23 (29,1%)	23 (27,7%)		Ref		
Desempregado	7 (8,9%)	18 (21,7%)	0,077	2,57	[0,90;7,32]	0,944
Reformado/Doente	49 (62%)	42 (50,6%)	0,670	0,86	[0,42;1,74]	-0,154
Estado Civil, n (%)						
Solteiro	3 (3,8%)	7 (8,4%)		Ref		
Casado/União Facto	61 (77,2%)	64 (77,1%)	0,262	0,45	[0,11;1,82]	-0,799
Divorciado/Separado	5 (6,3%)	6 (7,2%)	0,469	0,51	[0,09;3,11]	-0,665
Viúvo	10 (12,7%)	6 (7,2%)	0,115	0,26	[0,05;1,39]	-1,358
Anos de escolaridade completos, média \pm dp	7 \pm 4,37	7,39 \pm 3,74	0,544	1,02	[0,95;1,11]	0,024
Etnia, n (%)						
Etnia Branca	62 (78,5%)	57 (68,7%)		Ref		
Etnia Negra	17 (21,5%)	26 (31,3%)	0,160	1,66	[0,82;3,38]	0,509
Vive com quantas pessoas, média \pm dp	1,77 \pm 1,29	1,60 \pm 1,08	0,374	0,89	[0,69;1,15]	-0,116
Dificuldade financeira para comprar medicação, n (%)						
Sem dificuldade	40 (50,6%)	34 (41%)		Ref		
Alguma dificuldade	25 (31,6%)	32 (38,6%)	0,248	1,51	[0,75;3,02]	0,409
Muita dificuldade	14 (17,7%)	17 (20,5%)	0,407	1,43	[0,62;3,32]	0,357
Idade de diagnóstico de HTA, média \pm dp	49,37 \pm 12,14	47,82 \pm 12,37	0,400	0,99	[0,96;1,02]	-0,011
Número de medicamentos, média \pm dp	4,85 \pm 2,94	4,90 \pm 2,90	0,901	1,01	[0,90;1,12]	-0,011
Número medicamentos para HTA, Média \pm dp	1,58 \pm 0,89	1,88 \pm 0,98	0,039	1,47	[1,02;2,11]	0,383
IMC, n (%)						
Baixo peso/normal	12 (15,2%)	11 (13,3%)		Ref		
Excesso de peso	29 (36,7%)	29 (34,9%)	0,860	1,09	[0,41;2,87]	0,087
Obesidade	38 (48,1%)	43 (51,8%)	0,656	1,23	[0,49;3,12]	0,211
Tem diabetes (n%)						
não	57 (72,2%)	65 (78,3%)		Ref		
sim	22 (27,8%)	18 (21,7%)	0,364	0,72	[0,35;1,47]	-0,332
Tem colesterol (n%)						
sim	39 (49,4%)	43 (51,8%)		Ref		
não	40 (50,6%)	40 (48,2%)	0,756	0,91	[0,49;1,68]	-0,098
Risco cardiovascular (n%)						
sem	21 (26,6%)	22 (26,5%)		Ref		
com	58 (73,4%)	61 (73,5%)	0,990	1	[0,50;2,02]	0,004
Considera HTA controlada (n%)						
sim	68 (86,1%)	67 (80,7%)		Ref		
não	11 (13,9%)	16 (19,3%)	0,363	1,48	[0,64;3,41]	0,390

Tabela 26. Dados relativos às regressões logísticas multinomiais simples com variável dependente Mudança de adesão: comparação 0-Sempre Aderente vs 1-Melhor comportamento. As variáveis a entrar no modelo têm valor $p \leq 0,15$ (cont.).

Variável	Aderentes n=75	Não aderentes n=83	Valor p	OR	IC OR (95%)	β
Tem ajuda para controlar HTA (%) sim não	7 (8,9%) 72 (91%)	11 (13,3%) 72 (86,7%)	0,377	Ref 0,64	[0,23;1,73]	-0,452
Frequência da medição HTA (%) Medição regular Medição pouco regular Outras	55 (69,6%) 18 (22,8%) 6 (7,6%)	63 (75,9%) 14 (16,9%) 6 (7,2%)	0,264 0,856	Ref 0,66 0,86	[0,31;1,38] [0,17;4,43]	-0,423 -0,152
Recorreu ao médico de família no último ano (%) sim não	78 (98,7%) 1 (1,3%)	82 (98,8%) 1 (1,2%)	0,972	Ref 0,95	[0,06;15,47]	-0,050
Recorreu ao médico especialista no último ano (%) sim não	3 (3,8%) 76 (96,2%)	10 (12%) 73 (88%)	0,067	Ref 0,29	[0,08;1,09]	-1,244
Recorreu ao médico privado no último ano (%) sim não	4 (5,1%) 75 (94,9%)	9 (10,8%) 74 (89,2%)	0,186	Ref 0,44	[0,13;1,49]	-0,824
Recorreu ao serviço de urgência no último ano (%) sim não	6 (7,6%) 73 (92,4%)	6 (7,2%) 77 (92,8%)	0,929	Ref 1,05	[0,33;3,42]	0,053
Recorreu ao atendimento no centro de saúde no último ano (%) sim não	4 (5,1%) 75 (94,9%)	6 (7,2%) 77 (92,8%)	0,569	Ref 0,68	[0,19;2,52]	-0,379
Recorreu ao farmacêutico no último ano (%) sim não	15 (19%) 64 (81%)	19 (22,9%) 64 (77,1%)	0,542	Ref 0,79	[0,37;1,69]	-0,236
Foi seguido regularmente na farmácia no último ano (%) sim não	12 (15,2%) 67 (84,8%)	12 (14,5%) 71 (85,5%)	0,896	Ref 1,06	[0,45;2,52]	0,058
Recorreu ao enfermeiro no último ano (%) sim não	2 (2,5%) 77 (97,5)	3 (3,6%) 80 (96,4%)	0,692	Ref 0,69	[0,11;4,26]	-0,367
Controlo da TA 140/90 (%) controlado não controlado	38 (48,1%) 41 (51,9%)	45 (54,2%) 38 (45,8%)	0,437	Ref 0,78	[0,42;1,45]	-0,245
Média pressão arterial sistólica, média \pm dp	142,06 \pm 20,41	136,57 \pm 17,31	0,070	0,99	[0,97;1,00]	-0,015
Média pressão arterial diastólica, média \pm dp	83,68 \pm 12,10	82,81 \pm 10,41	0,610	0,99	[0,96;1,02]	-0,007

As variáveis a incluir no modelo são: Ocupação, Estado Civil, Número de medicamentos para HTA, Recorreu ao médico especialista no último ano e Média da pressão arterial sistólica.

Tabela 27. Resultados da regressão logística multinomial com variável dependente Mudança de adesão, comparação 0-Sempre Aderente vs 1-Melhor comportamento de adesão utilizando o método empírico.

Variáveis	β	Valor p	OR
Ocupação			
Trabalhador/Estudante			Ref
Desempregado	0,954	0,081	2,60
Reformado/Doente	-0,060	0,875	0,94
Estado Civil			
Solteiro			
Casado/União	-0,618	0,403	0,54
Divorciado/Separado	-0,704	0,457	0,49
Viúvo	-1,346	0,137	0,26
Número de Medicamentos	0,255	0,195	1,29
Recorreu ao médico especialista no último ano			
Sim			Ref
Não	-1,135	0,111	0,32
Média da pressão arterial sistólica.	-0,014	0,105	0,99

A única variável significativa para o modelo é Ocupação (Desempregado). (valor $p < 0,10$)

Tabela 28. Dados relativos às regressões logísticas multinomiais simples com variável dependente Mudança de adesão: comparação 0-Sempre Aderente vs 2-Mau comportamento. As variáveis a entrar no modelo têm valor $p \leq 0,15$.

Variável	Aderentes n=79	Não aderentes n=123	Valor p	OR	IC OR (95%)	β
Idade, média \pm dp	63,35 \pm 9,64	57,89 \pm 10,84	<0,001	0,95	[0,92;0,98]	-0,053
Género, n (%)						
Feminino	41 (51,9%)	70 (56,9%)		Ref		
Masculino	38 (48,1%)	53 (43,1%)	0,485	0,82	[0,46;1,44]	-0,202
Ocupação, n (%)						
Trabalhador/Estudante	23 (29,1%)	66 (53,7%)		Ref		
Desempregado	7 (8,9%)	16 (13%)	0,658	0,80	[0,29;2,18]	-0,228
Reformado/Doente	49 (62%)	41 (33,3%)	<0,001	0,29	[0,16;0,55]	-1,232
Estado Civil, n (%)						
Solteiro	3 (3,8%)	18 (14,6%)		Ref		
Casado/União Facto	61 (77,2%)	86 (69,9%)	0,025	0,24	[0,07;0,83]	-1,448
Divorciado/Separado	5 (6,3%)	13 (10,6%)	0,305	0,43	[0,09;2,14]	-0,836
Viúvo	10 (12,7%)	6 (4,9%)	0,004	0,10	[0,02;0,49]	-2,303
Anos de escolaridade completos, média \pm dp	7 \pm 4,37	6,41 \pm 3,83	0,291	0,96	[0,89;1,03]	-0,039
Etnia, n (%)						
Etnia Branca	62 (78,5%)	45 (36,6%)		Ref		
Etnia Negra	17 (21,5%)	78 (63,4%)	<0,001	6,32	[3,30;12,11]	1,844
Vive com quantas pessoas, média \pm dp	1,77 \pm 1,29	2,66 \pm 1,74	<0,001	1,51	[1,21;1,87]	0,411
Dificuldade financeira para comprar medicação, n (%)						
Sem dificuldade	40 (50,6%)	46 (37,4%)		Ref		
Alguma dificuldade	25 (31,6%)	38 (30,9%)	0,407	1,32	[0,68;2,55]	0,279
Muita dificuldade	14 (17,7%)	39 (31,7%)	0,020	2,42	[1,15;5,09]	0,885
Idade de diagnóstico de HTA, média \pm dp	49,37 \pm 12,14	44,97 \pm 11,44	0,012	0,97	[0,95;0,99]	-0,031
Número de medicamentos, média \pm dp	4,85 \pm 2,94	4,57 \pm 2,58	0,479	0,96	[0,87;1,07]	-0,037
Número medicamentos para HTA, Média \pm dp	1,58 \pm 0,89	1,82 \pm 0,92	0,067	1,38	[0,98;1,95]	0,322
IMC, n (%)						
Baixo peso/normal	12 (15,2%)	13 (10,6%)		Ref		
Excesso de peso	29 (36,7%)	52 (42,3%)	0,276	1,66	[0,67;4,10]	0,504
Obesidade	38 (48,1%)	58 (47,2%)	0,448	1,41	[0,58;3,41]	0,343
Tem diabetes (n%)						
não	57 (72,2%)	85 (69,1%)		Ref		
sim	22 (27,8%)	38 (30,9%)	0,644	1,16	[0,62;2,16]	0,147
Tem colesterol (n%)						
sim	39 (49,4%)	61 (49,6%)		Ref		
não	40 (50,6%)	62 (50,4%)	0,975	0,99	[0,56;1,74]	-0,009
Risco cardiovascular (n%)						
sem	21 (26,6%)	31 (25,2%)		Ref		
com	58 (73,4%)	92 (74,8%)	0,827	1,07	[0,56;2,05]	0,072
Considera HTA controlada (n%)						
sim	68 (86,1%)	86 (69,9%)		Ref		
não	11 (13,9%)	37 (30,1%)	0,010	2,66	[1,26;5,60]	0,978

Tabela 28. Dados relativos às regressões logísticas multinomiais simples com variável dependente Mudança de adesão: comparação 0-Sempre Aderente vs 2-Mau comportamento. As variáveis a entrar no modelo têm valor $p \leq 0,15$ (cont.).

Variável	Aderentes n=79	Não aderentes n=123	Valor p	OR	IC OR (95%)	β
Tem ajuda para controlar HTA (%) sim não	7 (8,9%) 72 (91%)	22 (17,9%) 101 (82,1%)	0,080	Ref 0,45	[0,18;1,10]	-0,807
Frequência da medição HTA (%) Medição regular Medição pouco regular Outras	55 (69,6%) 18 (22,8%) 6 (7,6%)	88 (71,5%) 18 (14,6%) 17 (13,8%)	0,435 0,772	Ref 0,77 0,80	[0,40;1,49] [0,17;3,70]	-0,263 -0,227
Recorreu ao médico de família no último ano (%) sim não	78 (98,7%) 1 (1,3%)	121 (98,4%) 2 (1,6%)	0,837	Ref 1,29	[0,11;14,46]	0,254
Recorreu ao médico especialista no último ano (%) sim não	3 (3,8%) 76 (96,2%)	12 (9,8%) 111 (90,2%)	0,128	Ref 0,36	[0,10;1,34]	-1,008
Recorreu ao médico privado no último ano (%) sim não	4 (5,1%) 75 (94,9%)	10 (8,1%) 113 (91,9%)	0,407	Ref 0,60	[0,18;1,99]	-0,506
Recorreu ao serviço de urgência no último ano (%) sim não	6 (7,6%) 73 (92,4%)	17 (13,8%) 106 (86,2%)	0,180	Ref 0,51	[0,19;1,36]	-0,668
Recorreu ao atendimento no centro de saúde no último ano (%) sim não	4 (5,1%) 75 (94,9%)	9 (7,3%) 114 (92,7%)	0,526	Ref 0,68	[0,20;2,27]	-0,392
Recorreu ao farmacêutico no último ano (%) sim não	15 (19%) 64 (81%)	58 (47,2%) 65 (52,8%)	<0,001	Ref 0,26	[0,14;0,51]	-1,337
Foi seguido regularmente na farmácia no último ano (%) sim não	12 (15,2%) 67 (84,8%)	21 (17,1%) 102 (82,9%)	0,724	Ref 0,87	[0,40;1,89]	-0,139
Recorreu ao enfermeiro no último ano (%) sim não	2 (2,5%) 77 (97,5)	18 (14,6%) 105 (85,4%)	0,013	Ref 0,15	[0,03;0,67]	-1,887
Controlo da TA 140/90 (%) controlado não controlado	38 (48,1%) 41 (51,9%)	49 (39,8%) 74 (60,2%)	0,248	Ref 1,40	[0,79;2,48]	0,336
Média pressão arterial sistólica, média \pm dp	142,06 \pm 20,41	141,62 \pm 20,37	0,876	1,00	[0,98;1,01]	-0,001
Média pressão arterial diastólica, média \pm dp	83,68 \pm 12,10	88,45 \pm 10,77	0,004	1,04	[1,01;1,07]	0,039

As variáveis a incluir no modelo são: Idade, Ocupação, Estado Civil, Etnia, Vive com quantas pessoas, Dificuldade financeira para comprar medicação, Idade de diagnóstico HTA, Número de medicamentos para HTA, Considera HTA controlada, Tem ajuda para controlar HTA, Recorreu ao médico especialista no último ano, Recorreu ao farmacêutico no último ano, Recorreu ao enfermeiro no último ano, Média da pressão arterial diastólica.

Tabela29. Resultados da regressão logística multinomial com variável dependente Mudança de adesão, comparação 0-Sempre Aderente vs 2-Mudança/manutenção de não adesão utilizando o método empírico.

Variáveis	β	Valor p	OR
Idade	-0,008	0,776	0,99
Ocupação			Ref
Trabalhador/Estudante			
Desempregado	-0,386	0,500	0,68
Reformado/Doente	-0,619	0,219	0,54
Estado Civil			Ref
Solteiro			
Casado/União	-0,024	0,975	0,98
Divorciado/Separado	1,170	0,219	3,22
Viúvo	-0,693	0,461	0,50
Etnia			Ref
Branca			
Negra	1,447	0,002	4,25
Vive com quantas pessoas	0,043	0,773	1,04
Dificuldade financeira para comprar medicação			Ref
Sem dificuldade			
Alguma dificuldade	-0,606	0,145	0,55
Muita dificuldade	-0,363	0,458	0,70
Idade de diagnóstico HTA	0,004	0,842	1
Número de medicamentos para HTA	0,301	0,154	1,35
Considera HTA controlada			Ref
Sim			
Não	0,610	0,189	1,84
Tem ajuda para controlar HTA			Ref
Sim			
Não	-0,927	0,077	0,40
Recorreu ao médico especialista no último ano			Ref
Sim			
Não	-0,910	0,229	0,40
Recorreu ao farmacêutico no último ano			Ref
Sim			
Não	-0,962	0,018	0,38
Recorreu ao enfermeiro no último ano			Ref
Sim			
Não	-1,605	0,057	0,20
Média da pressão arterial diastólica	0,022	0,208	1,02

As variáveis significativas para o modelo são: Etnia, Tem ajuda para controlar HTA, Recorreu ao farmacêutico no último ano e Recorreu ao enfermeiro no último ano. (valor $p < 0,10$)

Para a regressão logística multinomial obtiveram-se duas comparações diferentes: A primeira entre 0-Sempre Aderente e 1-Melhor comportamento de adesão e a segunda entre 0-Sempre Aderente e 2-Mudança/manutenção de não adesão.

Comecemos por observar a tabela 26 com os resultados da regressão multinomial simples (bivariada) para a comparação entre 0-Sempre aderente e 1-Melhor comportamento de adesão:

Podemos verificar que Indivíduos desempregados têm mais do dobro de hipótese de passar a um melhor comportamento de adesão comparativamente aos Trabalhadores/estudantes/donas de casa; reformados/doentes/permanentemente incapacitados têm menos 12% de hipótese de melhorar o comportamento de adesão; indivíduos viúvos têm menos 74% de hipótese de melhorar o comportamento de adesão comparativamente aos solteiros; Quanto mais medicamentos para a hipertensão arterial maior a hipótese de melhorar o comportamento de adesão; Indivíduos que não recorreram ao especialista no último ano têm menor hipótese de melhorar o seu comportamento de adesão (menos 71% de hipótese) comparativamente aos que consultaram o especialista.

Observando esta mesma tabela, as variáveis a entrar no modelo múltiplo de comparação 0-Sempre aderente e 1-Melhor comportamento de adesão são: Ocupação, Estado Civil, Número de medicamentos para HTA, Recorreu ao médico especialista no último ano e Média da pressão arterial sistólica. A tabela 27 indica-nos o resultado da regressão múltipla: A única variável importante no modelo de regressão logística multinomial multivariada é **Ocupação**, com os desempregados com mais do dobro das hipóteses de ter um melhor comportamento de adesão comparando com os trabalhadores/estudantes/donas de casa.

Observemos agora a tabela 28 a qual apresenta os resultados da regressão multinomial simples (bivariada) para a comparação entre 0-Sempre aderente e 2-Mudança/manutenção de não adesão: A etnia negra tem seis vezes mais hipótese de ter uma mudança/manutenção de não adesão do que a etnia branca; quanto maior o número de pessoas em casa maior a hipótese de ter uma mudança/manutenção de não adesão (OR = 1.04); indivíduos com muita dificuldade financeira para comprar os medicamentos para HTA têm mais do dobro das hipóteses de ter uma mudança/manutenção de não adesão comparando com indivíduos sem dificuldade para comprar os medicamentos; indivíduos que dizem não ter a HTA controlada têm mais do dobro das hipóteses de ter uma mudança/manutenção de não adesão do que indivíduos que consideram que têm a HTA controlada. Verifica-se ainda que pessoas com mais idade têm menos hipótese de ter uma mudança/ manutenção de não adesão; indivíduos reformados/doentes/permanentemente incapacitados têm menos

71% de hipótese de ter uma mudança/manutenção de não adesão comparativamente com os trabalhadores/estudantes/donas de casa; Casados têm menos 76% de hipótese de ter uma mudança/manutenção de não adesão comparando com os solteiros, o mesmo acontece com os viúvos com menos 90% de hipótese de ter uma mudança/manutenção de não adesão que os solteiros; o aumento da idade de diagnóstico da HTA diminui as hipóteses de ter uma mudança/manutenção de não adesão; indivíduos sem ajuda para controlar a HTA têm menos 55% de hipóteses de ter uma mudança/manutenção de não adesão, comparando com os que necessitam de ajuda para manter a Hipertensão controlada; Indivíduos que não recorreram ao especialista no último ano têm menos 64% de hipóteses de ter uma mudança/manutenção de não adesão, indivíduos que não recorreram ao farmacêutico no último ano têm menos 74% de hipótese de ter uma mudança/manutenção de não adesão; o mesmo acontece aos que não recorreram ao enfermeiro no último ano, menos 85% de hipótese de ter uma mudança/manutenção de não adesão.

As variáveis a entrar no modelo múltiplo são: Idade, Ocupação, Estado Civil, Etnia, Vive com quantas pessoas, Dificuldade financeira para comprar medicação, Idade de diagnóstico HTA, Número de medicamentos para HTA, Considera HTA controlada, Tem ajuda para controlar HTA, Recorreu ao médico especialista no último ano, Recorreu ao farmacêutico no último ano, Recorreu ao enfermeiro no último ano, Média da pressão arterial diastólica.

Observando a tabela 29 com o resultado da regressão multinomial multivariada para a comparação entre 0-Sempre aderente e 2-Mudança/manutenção de não adesão, temos:

As variáveis importantes no modelo são: **Etnia, Tem ajuda para controlar HTA, Recorreu ao farmacêutico no último ano, Recorreu ao enfermeiro no último ano**, com o facto de ter ajuda para controlar HTA, ter ido ao farmacêutico no último ano e ter consultado o enfermeiro como fatores indicadores de mudança/manutenção de não adesão e a Etnia negra 4 vezes mais hipóteses de apresentar mudança/manutenção de não adesão que a etnia branca.

5.4 Outros estudos

Foram ainda analisadas duas regressão logística simples, com as variáveis resposta “Classificação” e “Mudanças” (duas variáveis) sendo a variável independente a variável que determina se o paciente é seguido no centro de saúde ou numa unidade de saúde familiar. O objetivo será verificar se existe alguma relação entre as variáveis.

A codificação da variável CS_USF_cod é a seguinte: 0-Consultado no centro de saúde; 1-Consultado na unidade de saúde familiar.

Tabela 30. Resultado da regressão logística com variável resposta Classificacao

	β	Valor p
Intercept	0.390	0.015
CS_USF_cod Centro de saúde USF	0.002	0.993

Tabela 31. Resultado da regressão logística com variável resposta Mud_adesao com 2 categorias

	β	Valor p
Intercept	-0.416	0.010
CS_USF_cod Centro de saúde USF	0.319	0.186

Tabela 32. Resultado da regressão multinomial com variável resposta Mud_adesao, comparação y=0 vs y=1

		β	Valor p
Comparação y=0 vs y=1	Intercept	0.021	0.919
	CS_USF_cod Centro de saúde USF	0.072	0.823
Comparação y=0 vs y=2	Intercept	0.288	0.132
	CS_USF_cod Centro de saúde USF	0.356	0.224

Em nenhuma das regressões logísticas efetuadas a variável CS_USF_cod se mostrou significativa. Podemos assim concluir que não existem diferenças entre consultar o Centro de saúde ou a unidade de saúde familiar para qualquer uma das variáveis resposta utilizadas (Classificação, Mudança de adesão com 2 categorias e Mudança de adesão com 3 categorias).

Capítulo 6 Discussão/ Conclusão

A utilização de modelos de regressão logística neste estudo permite-nos verificar quais os fatores relacionados com a variável Classificação, que indica se um doente é aderente ou não à medicação no momento inicial do estudo, e Mudança de adesão, que indica se um doente altera ou não a sua adesão à medicação num intervalo de tempo de seis meses, num conjunto de variáveis aparentemente significativas. A vantagem destes modelos relativamente a regressões simples entre a variável resposta e cada uma das variáveis é a obtenção de variáveis importantes para a classificação e mudança de adesão, relacionadas com a variável dependente dentro de um conjunto e não relacionadas uma a uma. Foram efetuados dois tipos de modelo de regressão logística múltipla: para o primeiro tipo utilizou-se o ‘método empírico’ para obtenção das variáveis a incluir no modelo de regressão logística múltipla, e para o segundo tipo de modelo utilizou-se o ‘método stepwise’ para o mesmo fim. O ‘método empírico’ para obter as variáveis a utilizar no modelo é utilizado em diversos estudos clínicos, como por exemplo o projeto PRAdA (Peixoto et al); já o ‘método stepwise’ é o método considerado mais eficiente para a escolha das variáveis a constar no modelo final. O modelo obtido através do ‘método stepwise’ demonstrou, nos dois casos, um ajustamento ligeiramente melhor ou semelhante de acordo com os valores obtidos pela curva ROC. Foi ainda efetuado um modelo de regressão multinomial que permitiu perceber quais as variáveis mais associadas a melhores comportamento de adesão. Os modelos de regressão logística binários e multinomiais são boas ferramentas de determinação de fatores associados com variáveis resposta mas os seus resultados devem ser interpretados com cuidado. Bons resultados dos modelos de regressão não dependem apenas do seu uso adequado mas também a forma como foram recolhidos os dados.

Este estudo tem algumas limitações que devem ser indicadas. Outros potenciais determinantes tais como a comunicação entre o prestador de cuidados e o paciente, depressões, perceção do paciente sobre a doença e crenças em tratamentos que também poderiam estar associados com a não adesão/mau comportamento de adesão não puderam ser explorados neste estudo. Este estudo foi conduzido apenas em Lisboa e Vale do Tejo e os resultados não podem ser extrapolados para outras regiões de Portugal. Todos os resultados são referentes a esta amostra específica de 424/285 indivíduos seguidos no momento Basal/6 meses. Para outra amostra diferente os resultados poderiam ter sido diferentes logo estes resultados terão de ser tratados com algum cuidado de forma a não levar a conclusões erradas.

Através da análise de resultados verificamos que a percentagem de não aderentes diminui com o tempo (15% menos em 6 meses, 21,8% em 12 meses) assim

como os não aderentes frequentes que diminuem em 6 meses mas diminuem muito mais em 12 meses de seguimento. Os sempre não aderentes apresentam a maior percentagem (nas mudanças de adesão) nos dois seguimentos, 6 meses e 12 meses, sendo um pouco mais pequena aos 12 meses. Estas observações poder-nos-ão levar a concluir que a não adesão diminui com o passar do tempo e com a aceitação da doença e dos problemas associados. Esta conclusão terá de ser pensada com cuidado pois a base de dados Basal-12meses é um pouco mais pequena, com alguns indivíduos diferentes da utilizada para Basal-6meses e também foram encontrados mais erros/problemas nas respostas. Deveremos ter em consideração que muitos indivíduos desistiram de participar a meio do estudo e vários dos que não o fizeram responderam pior/com mais erros/com menos paciência e vontade, o que é natural ao fim de um ano de estudo. Também o facto de ser um questionário de auto resposta poderá ter inflacionado os resultados do estudo porque os respondentes têm tendência a responder o que é socialmente aceitável podendo não assumir a não adesão (Dijkstra et al).

Efetuando o estudo de regressão logística múltipla com resposta “Classificação” através do “método empírico” verifica-se que todos os estados civis são mais propícios à adesão (Casado, Divorciado, Viúvo) comparando com os Solteiros, isto é, os solteiros têm maior propensão para a não adesão. Já em estudos anteriores foi verificado que indivíduos casados são menos aderentes à medicação (Morris AB, Li J et al). Também os indivíduos que necessitam de ajuda para controlar a sua hipertensão arterial são mais não aderentes. É possível que indivíduos casados/divorciados/viúvos sintam mais necessidade de se cuidar de forma a não deixarem a família próxima (esposa, filhos,...) em cuidados com a sua saúde e que pessoas mais independentes no seu tratamento obtenham melhores resultados no controlo da sua hipertensão não tendo assim tendência para não aderir à medicação. Já através do “método Stepwise” as variáveis importantes no modelo foram bastante diferentes: Pessoas mais novas, de sexo masculino, de etnia negra, que consideram a HTA não controlada e com ajuda para controlar HTA são menos aderentes, como já foi observado em estudos anteriores, efetuados em todo o mundo (Bosworth HB et al)(Ramil A; Ahmad MS et al)(Hashimi S; Afridi M et al). O modelo efetuado através do “método Stepwise” obteve um maior valor de AUC que o modelo empírico, isto é, mostra estar ligeiramente melhor ajustado.

Efetuando o estudo de regressão logística multinomial com resposta “Mudança de adesão” com duas categorias (0–Mudança positiva vs 1-Mudança negativa), através do “método empírico” verificamos que ser desempregado, com dificuldades para comprar os medicamentos, sem ajuda para controlar a HTA e sem consultar o farmacêutico e o enfermeiro no ultimo ano são fatores que incitam à mudança positiva. Já o facto de ser divorciado aumenta (e muito) a probabilidade de ter uma mudança negativa, assim como a etnia negra tem mais tendência para ter mudança

negativa. Também o aumento do número de pessoas em casa propicia as mudanças negativas. Quanto à ocupação, é possível que indivíduos com mais tempo disponível tenham mais tempo para cumprir a medicação. Já os indivíduos com dificuldade para comprar medicamentos poderão ter mais cuidado em cumprir a medicação tendo em conta o esforço que fizeram para a comprar. Indivíduos que consultam o farmacêutico e o enfermeiro são mais favoráveis a ter complicações com a paragem da medicação, a ter reações adversas, etc. Logo será natural que estejam mais relacionados com a mudança negativa na adesão. Indivíduos recentemente divorciados poderão ter mudanças negativas na adesão por perderem uma ajuda importante, a do cônjuge. A etnia negra é considerada em diversos estudos como não aderente à medicação (Bosworth, HB et al); não é de estranhar que apresente os piores comportamentos de adesão, o aumento do número de pessoas em casa poderá indicar uma tendência para estratos familiares mais pobres e com menos condições que poderão levar a piores comportamentos de adesão. Este tipo de estudo foi ainda pouco explorado, existindo apenas um estudo semelhante (que é do nosso conhecimento) (Chunhua Ma et al). No entanto, em estudos semelhantes, o apoio familiar foi considerado importante para manter um bom comportamento de adesão, ao contrário do que se obteve com este estudo: pessoas sem ajuda para controlarem HTA mantêm um melhor comportamento de adesão. (Chunhua Ma et al). Através do “método stepwise” obtiveram-se resultados bastante semelhantes ao “método empírico”, com uma única variável acrescentada ao modelo: um maior nível de escolaridade como fator propiciador da mudança positiva. Este resultado é consistente com outros estudos desenvolvidos anteriormente (Chunhua Ma et al) e pode indicar que deverá ser dado um maior apoio a doentes com menos escolaridade.

Através do estudo de regressão logística multinomial multivariada para a comparação 0-Sempre bom comportamento de adesão vs 1-Melhor comportamento de adesão a única conclusão retirada é que indivíduos Desempregados têm tendência para melhorar o comportamento de adesão (novamente faz sentido a razão mencionada anteriormente, indivíduos com mais tempo poderão dedicar-se mais a tomar a medicação como deve ser). Para a comparação 0-Sempre bom comportamento de adesão vs 2-Manutenção de não adesão encontramos novamente a etnia negra relacionada com o mau comportamento de adesão, e indivíduos sem ajuda para controlar HTA que não consultaram o farmacêutico e não consultaram o enfermeiro no último ano mais relacionados com o bom comportamento de adesão, como já tinha sido verificado na regressão efetuada para mudança com duas categorias.

Uma potencial razão para os doentes hipertensos não cumprirem o tratamento e recusar as recomendações dos médicos poderá estar relacionada com a medicação que terá de ser tomada toda a vida (Wang, S Z). Outra causa poderá estar relacionada com o facto de a hipertensão arterial ser uma doença com poucos sintomas. (Giorgi D)

Assim, profissionais de saúde deverão ter especiais cuidados com os indicadores de maus comportamentos de adesão (ou sempre não aderente ou mudança de aderente para não aderente): divorciados, trabalhadores/estudantes (pouco tempo), sem dificuldades para comprar a medicação, com várias pessoas a morar na mesma casa (possível indicador de pobreza), Etnia negra, com idas frequentes ao farmacêutico e enfermeiro (indicativo de mais problemas relacionados com a medicação ou falta dela) e com ajuda para controlar a hipertensão (sem autonomia).

Capítulo 7 Bibliografia

- **Almeida, Helcia et al.** *Adesão a Tratamentos entre Idosos. Comunicação em Ciências da Saúde*. 18 (1). Brasília (Janeiro/Março 2007). Pp. 57-67.
- **Bosworth HB, Powers B, Grubber JM, Thorpe CT, Olsen MK, Orr M, Oddone EZ.** Racial differences in blood pressure control: potential explanatory factors. *J Gen Intern Med*. May 2008, Vol 23, 5, pp. 692-8.
- **Chunhua Ma, Shaoxian Chen, Ying Zhou, Chunfeng Huang.** Treatment adherence of Chinese patients with hypertension: A longitudinal study. *Applied Nursing Research* 26 (2013), 225-231.
- **David W. Hosmer and Stanley Lemeshow.** *Applied logistic regression*, second edition, 2000.
- **Denhaerynck K, Van Camp Y, Villa L, MacDonald K, Vancayzeele S, Brié H, Aerts A, Hermans C, Abraham I.** Hierarchical linear and logistic modeling of blood pressure outcomes in adherent versus non-adherent treated with Valsartan: pooled analysis of 13,221 patients, 2012.
- **Dias, A.M.; Cunha, M.; Santos, A.; Neves, A.; Pinto, A.; Silva, A.; Castro, S.** Adesão ao regime terapêutico na doença crónica: Revisão da literature. *Millenium*, 40: 201-219. 2011.
- **Dijkstra W, Smit JH, Comijs HC.** Using social desirability scales in research among the elderly. *Qual Quant* . 2001, Vol. 35, pp. 107–115.
- **Direcção geral de saúde.** Hipertensão arterial: definição e classificação, (2003).
- **Direcção Geral de saúde.** Diagnóstico, Tratamento e Controlo da Hipertensão Arterial, circular normativa (2004).
- **E. Márquez Contreras, J.L. Martín de Pablos, V. Gil Guillén, N. Martel Claros, J. Motero Carrasco, J.J. Casado Martínez en representación del grupo de trabajo sobre Cumplimiento de la Asociación de la Sociedad Española de Hipertensión y Liga Española para la Lucha contra la Hipertensión Arterial.** La inercia clínica profesional y el incumplimiento farmacológico: como influyen en el control de la hipertensión arterial? Estudio CUMAMPA. *Hipertensión (Madr.)*, 2008.
- **Emilio Márquez-Contreras, Mariano de la Figuera-Von Wichmann, Josep Franch-Nadal, Jose Luis Listerri-Caro, Vicente Gil-Guillén, José Luis Martín-de Pablos, José Joaquin Casado-Martínez, Nieves Martell-Claros.** Los pacientes com alto riesgo vascular toman correctamente la medicación antihipertensiva? Estudio Cumple-MEMS. *Rev. Esp. Cardiol.*, 2012.

- **Faria, Heloisa Turcatto Gimenes.** Factores Relacionados à Adesão do Paciente Diabético à Terapêutica Medicamentosa. Dissertação de Mestrado não publicada. Universidade de São Paulo. Escola de Enfermagem de Ribeirão Preto, 2008.
- **Garson, G.David.** Fundamentals of Hierarchical Linear and Multilevel Modeling.
- **Giorgi D.** Estratégias para melhorar a adesão ao tratamento anti-hipertensivo. Revista Brasileira Hipertens, vol.13(1):47-50, 2006.
- **Hashmi SK, Afridi MB, Abbas K, Sajwani RA, Saleheen D, Frossard PM, Ishaq M, Ambreen A, Ahmad U.** Factors associated with adherence to anti-hypertensive treatment in Pakistan. March 14, 2007.
- **Hasimur Rahaman Khan and J. Ewart H. Shaw.** Multilevel logistic regression analysis applied to binary contraceptive prevalence data. Journal of data science 9, 2011.
- **Haynes, RB et al.** Intervention for Enhancing Medication Adherence. The Cochrane Database of Systematic Reviews, 2008. Issue 2.
- **João Gomes, DEIO – FCUL.** Regressão binária. O modelo logístico, 2011.
- **Joseph, Lawrence.** Apontamentos Goodness of Fit in Logistic Regression. Division of Clinical Epidemiology, McGill University Health Centre, Royal Victoria Hospital.
- **Larissa Grigoryan, Valory N. Pavlik, David J. Hyman.** Patterns of nonadherence to antihypertensive therapy in primary care. February 2013, Vol. 15, No 2.
- **Leite, Silvana Nair & Vasconcellos, Maria da Penha.** Adesão à Terapêutica Medicamentosa: Elementos para a Discussão de Conceitos e Pressupostos Adotados na Literatura. Ciência e Saúde Coletiva, 2003, 8 (3). Pp. 775-782.
- **Macedo ME, Lima MJ, Silva AO, et al.** Prevalence, awareness, treatment and control of hypertension in Portugal: the PAP study. *J Hypertens*. Sep 2005 , Vol. 23, 9, pp. 1661-6.
- **Machado, Maria Manuela.** Adesão o Regime Terapêutico: Representações das Pessoas com IRC sobre o contributo dos Enfermeiros. Instituto de Educação e Psicologia. (2009) .
- **Morris AB, Li J, Kroenke K, Bruner-England TE, Young JM, Murray MD.** Factors associated with drug adherence and blood pressure control in patients with hypertension. *Pharmacotherapy*. Apr.2006, Vol.26, 4, pp. 483-92.
- **Organization, World Health.** Adherence to Long Term Therapies: Evidence for Action. Geneva: WHO, 2003.
- **Peixoto C, Carrilho G, Alarcão V, Guerra F, Simões R, Fernandes M, Nicola P, Guiomar S, Nogueira P, Rocha E.** Intervenção PRAdA: Efeitos no Peso Induzidos por Proposta de Alteração de Comportamento Alimentar em Hipertensos Medicados Nativos Portugueses e Imigrantes de Origem Africana. *Acta Med Port* 2014, Mai-Jun; 27(3): 331-341.
- **Pedro Aguiar, em colaboração com Catarina Silva, Filipa Negreiro e Vera Vicente.** Eurotrials Scientific Consultants. Ficha formativa de estatística. Qual a

importância da curva ROC (Receiver Operating Characteristic)? Gauss, nº 17, Junho 2012.

- **Polónia J, Martins L, Pinto F, Nazaré J.** Prevalence, awareness, treatment and control of hypertension and salt intake in Portugal: changes over a decade. The PHYSA study. *J Hypertens*, 2014 Jun;32(6):1211-21.
- **Ramli A, Ahmad NS, Paraidathathu T.** Medication adherence among hypertensive patients of primary health clinics in Malaysia. *Patient Prefer Adherence*, 2012, Vol 6, pp. 613-22.
- **Souto D, Fernandes M, Simões R, Alarcão V, Lopes E, Gomez V, Nicola PJ, Rocha E.** Adherence to antihypertensive medication and associated factors among African migrants and Portuguese natives, 2012.
- **Vermeire E et al.** Patients Adherence to Treatment: Three Decades of Research. *Journal of Clinical Pharmacy and Therapeutics*. 2001, 26, 331-342.
- **Wang, S. Z.** Health education in improving adherence patients with hypertension. *The Nursing Scientific Journal*, 2003, 18(2), 89–91.

Capítulo 8 Anexos

8.1 Códigos utilizados em R

Neste capítulo serão indicados os códigos utilizados no programa R de forma a obter cada um dos modelos de regressão utilizados.

8.1.1 Leitura e reconhecimento das variáveis no programa R

```
ficheiro.dados<-read.table(file="ficheiro.txt",sep="\t",header=T)
```

```
is.data.frame(ficheiro.dados)
```

```
head(ficheiro.dados,10)
```

```
ficheiro.dados$Genero<-factor(ficheiro.dados$Genero)
```

```
ficheiro.dados$EstadoCivil<-factor(ficheiro.dados$EstadoCivil)
```

```
ficheiro.dados$cOcupacao<-factor(ficheiro.dados$cOcupacao)
```

```
ficheiro.dados$cFinancDific<-factor(ficheiro.dados$cFinancDific)
```

```
ficheiro.dados$IMC_3categorias<-factor(ficheiro.dados$IMC_3categorias)
```

```
ficheiro.dados$DiagnDiabetes<-factor(ficheiro.dados$DiagnDiabetes)
```

```
ficheiro.dados$cColestrol<-factor(ficheiro.dados$cColestrol)
```

```
ficheiro.dados$RCV<-factor(ficheiro.dados$RCV)
```

```
ficheiro.dados$cCtrlHTA<-factor(ficheiro.dados$cCtrlHTA)
```

```
ficheiro.dados$cAjudaHTA<-factor(ficheiro.dados$cAjudaHTA)
```

```
ficheiro.dados$FreqCod<-factor(ficheiro.dados$FreqCod)
```

```
ficheiro.dados$PA_Control_2cat<-factor(ficheiro.dados$PA_Control_2cat)
```

```
ficheiro.dados$cRecMedFamiliaUltAno<-factor(ficheiro.dados$cRecMedFamiliaUltAno)
```

```
ficheiro.dados$cRecEspecialistaUltAno<-factor(ficheiro.dados$cRecEspecialistaUltAno)
```

```

ficheiro.dados$cRecMedPrivadoUltAno<-
factor(ficheiro.dados$cRecMedPrivadoUltAno)

ficheiro.dados$cRecUrgenciaUltAno<-factor(ficheiro.dados$cRecUrgenciaUltAno)

ficheiro.dados$cRecAtendCSUltAno<-factor(ficheiro.dados$cRecAtendCSUltAno)

ficheiro.dados$cRecFarmaceuticoUltAno<-
factor(ficheiro.dados$cRecFarmaceuticoUltAno)

ficheiro.dados$cSegFarmaciaUltAno<-factor(ficheiro.dados$cSegFarmaciaUltAno)

ficheiro.dados$cEnfermUlt<-factor(ficheiro.dados$cEnfermUlt)

ficheiro.dados$Classificacao<-factor(ficheiro.dados$Classificacao)

ficheiro.dados$Etnia<-factor(ficheiro.dados$Etnia)

ficheiro.dados$Mud_adesao<-factor(ficheiro.dados$Etnia)


head(ficheiro.dados,10)

str(ficheiro.dados)

```

8.1.2 Modelos Univariados

Modelos univariados

Os modelos univariados foram efetuados para Classificacao e Mud_adesao. Apenas coloco os de variável resposta Classificacao pois tudo o resto foi efetuado da mesma forma, é só adaptar a variável.

```

R1<-glm(Classificacao ~ Idade,family = binomial, ficheiro.dados)

R2<-glm(Classificacao ~ Genero,family = binomial, ficheiro.dados)

R3<-glm(Classificacao ~ EstadoCivil,family = binomial, ficheiro.dados)

R4<-glm(Classificacao ~ nAnosEscola,family = binomial, ficheiro.dados)

R5<-glm(Classificacao ~ Etnia,family = binomial, ficheiro.dados)

R6<-glm(Classificacao ~ nPessoasCasa,family = binomial, ficheiro.dados)

R7<-glm(Classificacao ~ cOcupacao,family = binomial, ficheiro.dados)

R8<-glm(Classificacao ~ cFinancDific,family = binomial, ficheiro.dados)

```

```

R9<-glm(Classificacao ~ IMC_3categorias,family = binomial, ficheiro.dados)
R10<-glm(Classificacao ~ DiagnDiabetes,family = binomial, ficheiro.dados)
R11<-glm(Classificacao ~ RCV,family = binomial, ficheiro.dados)
R12<-glm(Classificacao ~ cColestrol,family = binomial, ficheiro.dados)
R13<-glm(Classificacao ~ cCtrlHTA,family = binomial, ficheiro.dados)
R14<-glm(Classificacao ~ cAjudaHTA,family = binomial, ficheiro.dados)
R15<-glm(Classificacao ~ FreqCod,family = binomial, ficheiro.dados)
R16<-glm(Classificacao ~ PA_Control_2cat,family = binomial, ficheiro.dados)
R17<-glm(Classificacao ~ cRecMedFamiliaUltAno,family = binomial, ficheiro.dados)
R18<-glm(Classificacao ~ cRecEspecialistaUltAno,family = binomial, ficheiro.dados)
R19<-glm(Classificacao ~ cRecMedPrivadoUltAno,family = binomial, ficheiro.dados)
R20<-glm(Classificacao ~ cRecUrgenciaUltAno,family = binomial, ficheiro.dados)
R21<-glm(Classificacao ~ cRecAtendCSUltAno,family = binomial, ficheiro.dados)
R22<-glm(Classificacao ~ cRecFarmaceuticoUltAno,family = binomial, ficheiro.dados)
R23<-glm(Classificacao ~ cSegFarmaciaUltAno,family = binomial, ficheiro.dados)
R24<-glm(Classificacao ~ MediaPAD,family = binomial, ficheiro.dados)
R25<-glm(Classificacao ~ cEnfermUlt,family = binomial, ficheiro.dados)
R26<-glm(Classificacao ~ idadeHTA,family = binomial, ficheiro.dados)
R27<-glm(Classificacao ~ Num_Meds_HTA,family = binomial, ficheiro.dados)
R28<-glm(Classificacao ~ Num_Meds_basal,family = binomial, ficheiro.dados)
R29<-glm(Classificacao ~ MediaPAS,family = binomial, ficheiro.dados)

```

Extraí os p-values das regressões com uma variável explanatória, e escrevi-os num ficheiro de texto

```

pvalue1 <- summary(R1)$coef[2, "Pr(>|z|)"]
pvalue2 <- summary(R2)$coef[2, "Pr(>|z|)"]
pvalue3 <- summary(R3)$coef[2, "Pr(>|z|)"]
pvalue4 <- summary(R4)$coef[2, "Pr(>|z|)"]

```

```

pvalue5 <- summary(R5)$coef[2, "Pr(>|z|)"]
pvalue6 <- summary(R6)$coef[2, "Pr(>|z|)"]
pvalue7 <- summary(R7)$coef[2, "Pr(>|z|)"]
pvalue8 <- summary(R8)$coef[2, "Pr(>|z|)"]
pvalue9 <- summary(R9)$coef[2, "Pr(>|z|)"]
pvalue10<-summary(R10)$coef[2, "Pr(>|z|)"]
pvalue11<- summary(R11)$coef[2, "Pr(>|z|)"]
pvalue12<- summary(R12)$coef[2, "Pr(>|z|)"]
pvalue13<- summary(R13)$coef[2, "Pr(>|z|)"]
pvalue14<- summary(R14)$coef[2, "Pr(>|z|)"]
pvalue15<- summary(R15)$coef[2, "Pr(>|z|)"]
pvalue16<- summary(R16)$coef[2, "Pr(>|z|)"]
pvalue17<- summary(R17)$coef[2, "Pr(>|z|)"]
pvalue18<- summary(R18)$coef[2, "Pr(>|z|)"]
pvalue19<- summary(R19)$coef[2, "Pr(>|z|)"]
pvalue20<- summary(R20)$coef[2, "Pr(>|z|)"]
pvalue21<- summary(R21)$coef[2, "Pr(>|z|)"]
pvalue22<- summary(R22)$coef[2, "Pr(>|z|)"]
pvalue23<- summary(R23)$coef[2, "Pr(>|z|)"]
pvalue24<- summary(R24)$coef[2, "Pr(>|z|)"]
pvalue25<- summary(R25)$coef[2, "Pr(>|z|)"]
pvalue26<- summary(R26)$coef[2, "Pr(>|z|)"]
pvalue27<- summary(R27)$coef[2, "Pr(>|z|)"]
pvalue28<- summary(R28)$coef[2, "Pr(>|z|)"]
pvalue29<- summary(R29)$coef[2, "Pr(>|z|)"]

vector.p.values<-
as.vector(cbind(pvalue1,pvalue2,pvalue3,pvalue4,pvalue5,pvalue6,pvalue7,pvalue8,pv
alue9,pvalue10,pvalue11,pvalue12,pvalue13,pvalue14,pvalue15,pvalue16,pvalue17,pv

```

```
alue18,pvalue19,pvalue20,pvalue21,pvalue22,pvalue23,
pvalue24,pvalue25,pvalue26,pvalue27,pvalue28,pvalue29))
```

```
NomesVariaveis<-
as.vector(cbind(names(R1$coef[2]),names(R2$coef[2]),names(R3$coef[2]),names(R4$
coef[2]),names(R5$coef[2]),names(R6$coef[2]),names(R7$coef[2]),

names(R8$coef[2]),names(R9$coef[2]),names(R10$coef[2]),names(R11$coef[2]),name
s(R12$coef[2]),names(R13$coef[2]),names(R14$coef[2]),names(R15$coef[2]),

names(R16$coef[2]),names(R17$coef[2]),names(R18$coef[2]),names(R19$coef[2]),na
mes(R20$coef[2]),names(R21$coef[2]),names(R22$coef[2]),

names(R23$coef[2]), names(R24$coef[2]), names(R25$coef[2]), names(R26$coef[2]),
names(R27$coef[2]),names(R28$coef[2]),names(R29$coef[2])))

Matriz.NomesePvalues<-cbind(NomesVariaveis,vector.p.values)
```

```
write.table(Matriz.NomesePvalues,file="Resultados regressões
Univariadas.txt",sep="\t")
```

8.1.3 Modelo logístico com as variáveis escolhidas pelas regressões

Considerando que as variáveis escolhidas através das regressões anteriores foram Idade, cOcupacao e EstadoCivil poderemos obter o respetivo modelo logístico com variável resposta classificação da seguinte forma:

```
summary(glm (Classificacao ~ Idade+ cOcupacao+EstadoCivil, family = binomial,
ficheiro.dados))
```

8.1.4 Modelo logístico utilizando o método Stepwise

É necessário efetuar primeiro um modelo com todas as variáveis para que o método Stepwise possa escolher as variáveis a incluir no modelo.

```
ModeloGlobal<-glm (Classificacao ~ . , family = binomial, ficheiro.dados)
```

```
summary(ModeloGlobal)

Modelo.Final<-step(ModeloGlobal)

summary(Modelo.Final)
```

8.1.5 Modelo Multinomial

O reconhecimento das variáveis tem de ser efetuado para que o programa consiga trabalhar com as variáveis. Neste caso será efetuada a regressão multinomial com variável resposta Mud_adesao com três categorias. A categoria de referência foi a categoria 0.

```
library(mlogit)

ficheiro.dados$Mud_adesao<-relevel(ficheiro.dados$Mud_adesao,"0")

Dados.Modelo<-mlogit.data(ficheiro.dados, varying = NULL, shape = "wide",
choice = "Mud_adesao")

str(Dados.Modelo)

write.table(Dados.Modelo,file="dados.txt",sep="\t",append=F)

# Modelo multinomial

summary(mlogit(Mud_adesao ~ 1|Idade+ cOcupacao+ EstadoCivil,
data=Dados.Modelo,reflevel = "0"))
```